

Evolution modularer Multienzymsysteme des bakteriellen Sekundärstoffwechsels

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

im Fach Biologie

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von

Holger Michael Jenke-Kodama, geb. Jenke
geboren am 25.09.1969 in Reutlingen

Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Christoph Marksches

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I

Prof. Dr. Christian Limberg

Gutachter:

1. Prof. Dr. Thomas Börner
2. Prof. Dr. Rolf Müller
3. PD Dr. Hans von Döhren

Tag der mündlichen Prüfung: 19.10.2007

Inhaltsverzeichnis

Zusammenfassung.....	6
1 EINLEITUNG	7
1.1 SEKUNDÄRMETABOLISMUS	7
1.1.1 Sekundärmetabolismus – eine Begriffsbestimmung	7
1.1.2 Vorkommen von Sekundärmetaboliten in den Organismenreichen	8
1.2 POLYKETIDSYNTASEN.....	10
1.2.1 Fettsäuresynthasen	11
1.2.1.1 Biochemie der Fettsäuresynthese	12
1.2.1.2 Struktur der verschiedenen FAS-Typen.....	13
1.2.2 Polyketidsynthasen des Typs I.....	16
1.2.2.1 Modulare PKS-I	17
1.2.2.1.1 Modularität als Basis chemischer Variabilität	17
1.2.2.1.2 Beladungsmodule.....	19
1.2.2.1.3 Modulare PKS-I mit trans-AT-Aktivität	21
1.2.2.1.4 Abweichungen vom Grundschema der modularen PKS-I-Synthese.....	21
1.2.2.2 Iterative PKS-I.....	22
1.2.3 Polyketidsynthasen des Typs II.....	23
1.2.4 Polyketidsynthasen des Typs III	24
1.2.5 Terminologie der PKS	25
1.3 NICHTRIBOSOMALE PEPTIDSYNTASEN.....	25
1.3.1 Aufbau von NRPS und allgemeines Synthese-Schema.....	26
1.3.2 Enzymologie und Struktur der NRPS-Domänen.....	27
1.3.2.1 Adenylierungs-Domänen	27
1.3.2.2 Kondensations-Domänen und Cyclisierungs-Domänen	28
1.3.3 Selektion der Aminosäuren durch die A-Domänen.....	29
1.3.4 Modifizierende Enzymaktivitäten in NRPS.....	29
1.4 GEMISCHTE BIOSYNTHESE MIT PKS- UND NRPS-MODULEN	30
1.5 BIOTECHNOLOGISCHES POTENZIAL VON MODULAREN PKS-I UND NRPS	30
1.6 CYANOBAKTERIELLE SEKUNDÄRMETABOLITE DES PKS- UND NRPS-TYPS	31
1.6.1 Übersicht.....	31
1.6.2 Das Cyanobakterium <i>Nostoc punctiforme</i>	33
1.7 ZIELSTELLUNG.....	34
2 MATERIAL UND METHODEN	35
2.1 MATERIAL.....	35
2.1.1 Bakterienstämme	35

2.1.1.1	<i>Nostoc punctiforme</i>	35
2.1.1.2	<i>Escherichia coli</i>	35
2.1.2	Chemikalien	35
2.1.3	Radiochemikalien	36
2.1.4	Enzyme	36
2.1.5	Nucleinsäuren	36
2.1.6	Kits	37
2.1.7	Filter und Membrane	38
2.1.8	Sequenzdaten	38
2.2	METHODEN	38
2.2.1	Mikrobiologische Methoden	38
2.2.1.1	Kultivierungsbedingungen für Bakterien	38
2.2.1.2	Herstellung chemisch kompetenter <i>E. coli</i> -Zellen	39
2.2.1.3	<i>Triparental mating</i> zur Herstellung von <i>N. punctiforme</i> -Mutanten	39
2.2.2	Molekularbiologische und biochemische Methoden	40
2.2.2.1	Isolierung genomischer DNA aus <i>N. punctiforme</i>	40
2.2.2.2	Klonierungen	40
2.2.2.2.1	Allgemeine Methodik	40
2.2.2.2.2	pRL271-Konstrukt zur Herstellung einer <i>N. punctiforme</i> -Mutante	40
2.2.2.2.3	Klonierung der Expressionsvektoren für heterologe Expression	41
2.2.2.2.4	Klonierung der DNA-Fragmente für den <i>Electrophoretic mobility shift assay</i>	41
2.2.2.3	Reinigung von Proteinen mit His-Tag	42
2.2.2.4	Konzentrationsbestimmung von Protein-Lösungen	42
2.2.2.5	Protein-Gelelektrophorese und Western Blot	42
2.2.2.6	<i>Electrophoretic mobility shift assay</i> (EMSA)	43
2.2.2.7	ATP-Pyrophosphat-Austausch-Assay	44
2.2.2.8	Probengewinnung aus <i>N. punctiforme</i> -Kulturen	44
2.2.2.9	HPCL-Analytik	45
2.2.3	Bioinformatische Methoden	45
2.2.3.1	Datengewinnung	45
2.2.3.2	Domänen-Analyse	46
2.2.3.3	Sequenz-Alignments	46
2.2.3.4	Phylogenetische Analysen	47
2.2.3.4.1	Bayes-Verfahren	47
2.2.3.4.2	<i>Neighbour joining</i> -Methode	48
2.2.3.4.3	<i>Maximum Parsimony</i> -Analyse	48
2.2.3.4.4	<i>Maximum Likelihood</i> -Analyse	49
2.2.3.5	Selektionsanalyse	49
3	ERGEBNISSE	50
3.1	EVOLUTIONÄRE STELLUNG DER VERSCHIEDENEN PKS-TYPEN	50

3.1.1	Evolutionäre Zusammenhänge zwischen FAS und PKS	50
3.1.2	Phylogenetische Analyse von PKS-III aus verschiedenen Organismenreichen	57
3.2	EVOLUTIONÄRE ZUSAMMENHÄNGE ZWISCHEN BAKTERIELLEN MODULAREN PKS-I.....	59
3.2.1	Verteilung modularer PKS-I innerhalb der Bakteriengruppen	59
3.2.2	Zusammenhang zwischen der Zahl an PKS-I-Genen und der Genomgröße	62
3.2.3	Phylogenetische Analyse modularer PKS-I	63
3.2.3.1	Phylogenie der AT-Domänen.....	63
3.2.3.2	Phylogenie der KS-Domänen.....	66
3.2.3.3	Bestimmung von Genduplikationen, Genverlusten und horizontalem Gentransfer aus den phylogenetischen Bäumen.....	68
3.2.3.4	Vergleich der phylogenetischen Bäume der AT- und KS-Domänen	71
3.3	INTRAGENOMISCHE EVOLUTION MODULARER PKS-I	72
3.3.1	PKS-I-Cluster in <i>Streptomyces avermitilis</i>	73
3.3.2	Phylogenetische Einordnung in den Zusammenhang mit anderen PKS-I-Clustern aus Streptomyceten	75
3.3.3	Phylogenien der verschiedenen Domänen-Typen.....	77
3.3.4	Änderung der Substratspezifität durch Ersetzung von AT-Domänen.....	81
3.3.5	Veränderungen der Reduktionsstufe bewirkende Domänen-Ersetzungen	83
3.3.6	Selektion in den PKS-I-Clustern von <i>S. avermitilis</i>	87
3.4	UNTERSUCHUNG DER GENOM-SEQUENZ VON <i>NOSTOC PUNCTIFORME</i> AUF PKS- UND NRPS-CLUSTER	88
3.5	ANALYTISCHER NACHWEIS VON POLYKETID- UND PEPTID-SEKUNDÄRMETABOLITEN IN <i>NOSTOC PUNCTIFORME</i>	99
3.6	EXPERIMENTELLE CHARAKTERISIERUNG EINES PKS/NRPS-CLUSTERS AUS <i>NOSTOC PUNCTIFORME</i>	102
3.6.1	Charakterisierung des Transkriptionsregulators	103
3.6.1.1	Sequenz-Untersuchung und phylogenetische Einordnung des Transkriptionsregulators	103
3.6.1.2	Untersuchung der Cluster-Sequenz auf mögliche Bindungsorte des Regulators	105
3.6.1.3	Heterologe Expression des Regulators NpF3179	107
3.6.1.4	Untersuchung der DNA-Bindung des Regulators NpF3179 mittels EMSA	108
3.6.2	Charakterisierung des NRPS-Moduls NpF3173	109
3.6.2.1	Heterologe Expression des NRPS-Moduls NpF3173 in <i>Escherichia coli</i>	110
3.6.2.2	Adenylierungsspezifität des NRPS-Moduls NpF3173	111
3.6.3	Untersuchung einer Knockout-Mutation der NRPS-Komponente im Cluster VII	111
3.6.3.1	Überprüfung der Mutante NpF3173::Kan ^R	112
3.6.3.2	Auswirkungen der Knockout-Mutation	113
4	DISKUSSION.....	116
4.1	EVOLUTIONÄRE EINORDNUNG DER POLYKETIDSYNTASEN	116

4.1.1	Phylogenetische Zusammenhänge zwischen PKS-I, PKS-II und FAS.....	117
4.1.1.1	Phylogenetische Position der PKS-II	119
4.1.1.2	Verwandtschaft zwischen bakteriellen FAS-I und FAS aus Pilzen.....	119
4.1.1.3	Bakterielle iterative PKS-I und Glycolipid-Synthasen.....	122
4.1.1.4	Phylogenetische Position der Varianten modular aufgebauter PKS-I.....	123
4.1.2	Das System der eukaryotischen FAS und Zusammenhänge mit bakteriellen PKS des Typs I	124
4.1.2.1	Varianten eukaryotischer FAS und ihre phylogenetischen Positionen.....	125
4.1.2.2	Stammt die FAS-I der Tiere von bakteriellen PKS-I ab?	127
4.1.3	Evolutionäre Beziehungen zwischen den PKS-III aus unterschiedlichen Organismengruppen	131
4.2	EVOLUTION DER MODULAREN PKS-I INNERHALB DER BAKTERIEN	132
4.2.1	Verteilung der PKS-I in Bakterien	132
4.2.2	Muster in der Evolution von PKS des Typs I in Bakterien	135
4.2.2.1	Modell der Evolution von PKS-I in Bakterien	136
4.2.2.2	Einfluss von horizontalem Gentransfer auf die Evolution bakterieller PKS-I.....	137
4.2.2.3	Evolutionärer Vergleich von PKS-I-Genclustern aus Streptomyceten.....	138
4.2.3	Biokombinatorik in der intragenomischen Evolution der PKS-I.....	141
4.2.3.1	Vergleich der natürlichen Biokombinatorik mit experimentellen Methoden und Ergebnissen	141
4.2.3.2	Evolutionäre Bedeutung der modularen Enzymorganisation: Modularität und Evolvierbarkeit	144
4.3	PKS- UND NRPS-BIOSYNTHESE-CLUSTER IN <i>NOSTOC PUNCTIFORME</i>	147
4.3.1	Genome mining im Cyanobakterium <i>Nostoc punctiforme</i>	147
4.3.2	Isolierung eines möglichen Metaboliten des Biosynthese-Clusters VII	149
4.4	EVOLUTION VON BAKTERIELLEN PKS-I UND NRPS IM KONTEXT DER GENERIERUNG METABOLISCHER DIVERSITÄT	151
4.4.1	Generierung von Strukturdiversität.....	152
4.4.2	Hypothesen zur funktionellen Bedeutung von Sekundärmetaboliten	154
4.4.3	Welche Rolle spielen Selektion und Adaptation in der Evolution des Sekundärstoffwechsels?	157
5	LITERATUR	159
	Abkürzungsverzeichnis.....	177
	Danksagungen.....	179
	Publikationen.....	180
	Erklärung.....	181

Zusammenfassung

Polyketid-Synthasen des Typs I (PKS-I) und nichtribosomale Peptid-Synthetasen (NRPS) sind modular aufgebaute Multienzymsysteme, die zum bakteriellen Sekundärstoffwechsel gehören. Bei der an PKS-I ablaufenden Biosynthese werden ähnlich wie bei der Fettsäure-Biosynthese aus kurzen Carbonsäure-Einheiten schrittweise lange Kohlenstoffketten aufgebaut. Viele der produzierten Sekundärmetabolite besitzen interessante Bioaktivitäten, die medizinisch genutzt werden können. Zur biochemischen Charakterisierung dieser Biosynthese-Cluster und ihrer Metabolite existiert eine große Zahl wissenschaftlicher Studien, während über die Evolution dieser Enzymsysteme relativ wenig bekannt ist. In der vorliegenden Arbeit wurden die modular gebauten PKS des Typs I aus Bakterien auf verschiedenen Ebenen evolutionär eingeordnet und Mechanismen abgeleitet, die bei der evolutionären Ausformung dieser Systeme eine besondere Rolle spielen.

Im Gesamtzusammenhang mit biochemisch verwandten Systemen aus allen Organismengruppen zeigte sich, dass die PKS-I Teil eines langen Evolutionsprozesses sind, in dem sie eine zentrale Position einnehmen. Eine phylogenomische Untersuchung unter Einbeziehung aller modularen PKS-I, die in vollständig sequenzierten Genomen codiert werden, erlaubte ihre Verteilung in den verschiedenen Bakteriengruppen zu untersuchen und die spezifischen Beiträge von Genduplikationen, Genverlusten und horizontalem Gentransfer (HGT) zu ermitteln. Der Einfluss dieser Faktoren variiert stark zwischen den einzelnen Bakteriengruppen. Während sich in Actinobakterien und Cyanobakterien viele Proteine von gemeinsamen Vorläufern ableiten ließen, spielte HGT bei der Evolution der PKS-I in Proteobakterien eine sehr große Rolle. Anhand des Genoms des Actinobakteriums *Streptomyces avermilis* konnten die intragenomischen Evolutionsprozesse von modularen PKS-I rekonstruiert werden. Es zeigte sich, dass homologe Rekombination zwischen Modulen einen wesentlichen Anteil an der Umprogrammierung der Biosynthese-Prozesse hat und der wichtigste Mechanismus zur Generierung der Variabilität von Polyketid-Verbindungen ist.

Im Genom des Cyanobakteriums *Nostoc punctiforme* konnte durch *genome mining* eine Reihe von Biosynthese-Clustern identifiziert werden, die aus PKS- und NRPS-Komponenten aufgebaut sind. Die analytische Untersuchung von Proben des Zellextrakts und des Kulturüberstandes ermöglichte die Zuordnung zu bestimmten Metaboliten. Eines der Biosynthese-Cluster wurde hinsichtlich des produzierten Metaboliten und der Regulationsstruktur eingehender charakterisiert.

Die Ergebnisse wurden schließlich in den allgemeinen Zusammenhang der Evolution metabolischer Diversität gebracht, wie sie insbesondere für den durch seine Stoff- und Strukturvielfalt charakterisierten Sekundärmetabolismus von Bedeutung ist.

1 Einleitung

1.1 Sekundärmetabolismus

Der zelluläre Stoffwechsel zeichnet sich durch eine große Vielfalt chemischer Reaktionen und Verbindungen aus. Zur Systematisierung und Einteilung sind daher Begriffsbildungen üblich, mit denen Teilbereiche des Gesamtmetabolismus gegeneinander abgegrenzt werden. So lassen sich beispielsweise die Gesamtheit der Reaktionen, die der Bereitstellung von Energie aus geeigneten Verbindungen dienen, unter dem Begriff Energiestoffwechsel, die Syntheseprozesse als Aufbaustoffwechsel (Anabolismus) und die abbauenden Reaktionen als Abbaustoffwechsel (Katabolismus) zusammenfassen. Da der Stoffwechsel ein kompliziertes Netzwerk enzymkatalysierter chemischer Reaktionen darstellt, stoßen solche Abgrenzungsversuche im Allgemeinen auf die Schwierigkeit, dass sich die Begriffe teilweise überlappen. Viele Zwischenprodukte des Energiestoffwechsels sind z. B. gleichzeitig Ausgangsstoffe für Biosynthesen. Eine weitere Einteilungsmöglichkeit ist die Unterscheidung zwischen Primärstoffwechsel und Sekundärstoffwechsel. Diese Differenzierung ist für die vorliegende Arbeit von grundlegender Bedeutung, weshalb zunächst, ausgehend von der historischen Entwicklung, eine Begriffsbestimmung erfolgt. Anschließend wird ein nach den Organismenreichen geordneter Überblick über Stoffgruppen gegeben, die dem Sekundärstoffwechsel zuzurechnen sind.

1.1.1 Sekundärmetabolismus – eine Begriffsbestimmung

Bereits Mitte des 19. Jahrhunderts waren mehr als 2000 Naturprodukte bekannt (Liebig 1843). Die terminologische Unterscheidung zwischen primären und sekundären Stoffwechselprozessen wurde im Jahr 1891 von dem Pflanzenphysiologen Kossel in den wissenschaftlichen Diskurs eingeführt (Kossel 1891). Hintergrund war die Beobachtung, dass bestimmte Verbindungen in allen bislang untersuchten Zellen gefunden werden konnten, während sich andere als „die zufälligen oder für das Leben nicht unbedingt nötigen Zellstoffe“ klassifizieren ließen. Er schlug vor, die erste Gruppe, die er als wesentliche Bestandteile der Zelle ansah, als primär zu bezeichnen, die zweite Gruppe hingegen als sekundäre Stoffe abzugrenzen. Dies geschah in der Absicht, das Interesse der Forschung auf die als wichtiger erachteten ubiquitär vorkommenden Grundprozesse zu lenken.

In den folgenden Jahrzehnten etablierte sich der Begriff Sekundärmetabolismus zunehmend in der Literatur, wenngleich immer wieder Versuche unternommen wurden, die inhärente Unschärfe zu beseitigen oder den als unpassend empfundenen Terminus durch andere zu ersetzen, insbesondere weil die Klassifizierung als sekundär eine Abwertung gegenüber dem Grundstoffwechsel hinsichtlich der biologischen Bedeutung zu implizieren scheint. So schlug Zähner vor, besser von generellem und speziellem Stoffwechsel zu sprechen (Zähner 1979). Brattsten führte den Begriff „allochemicals“ ein, um zu betonen, dass viele pflanzliche Sekundärmetabolite eine Funktion in der Interaktion zwischen Organismen haben (Brattsten 1979).

Bis heute existiert keine befriedigende, alle Teilaspekte berücksichtigende Definition des Begriffs Sekundärmetabolismus. Praktikabel erscheint der Rückgriff auf den leichter abgrenzbaren Primärstoffwechsel, um dann im Sinne einer negativen Definition unter Sekundärstoffwechsel alle jene Umsetzungen zu subsumieren, welche die an den Primärstoffwechsel gestellten Kriterien nicht erfüllen. Als allgemein anerkannte Charakteristika des Primärstoffwechsels lassen sich nennen (Luckner 1990): (1) das ubiquitäre Vorkommen der beteiligten Enzyme und Verbindungen in den Organismengruppen, wobei insbesondere völlige Homogenität zwischen eng verwandten Taxa besteht, und (2) die grundlegende Bedeutung für die Erhaltung und Vermehrung des Organismus. Berücksichtigt man weiterhin die Gemeinsamkeiten bekannter Sekundärstoffwechselwege, so lassen sich Sekundärmetabolite in Anlehnung an die Darstellung Luckners (Luckner 1990) folgendermaßen bestimmen:

Sekundärmetabolite sind Verbindungen, die (1) taxonomisch inhomogen verteilt sind, so dass insbesondere auch eng verwandte Arten Unterschiede im Vorhandensein oder in der genauen Struktur aufweisen, (2) von spezifischen Enzymen hergestellt werden, (3) sich von Verbindungen des Primärstoffwechsels ableiten, wobei aus den Grundverbindungen sehr diverse und teilweise komplizierte chemische Strukturen entstehen, (4) keine direkt lebensnotwendige Funktion für die produzierende Zelle haben und schließlich (5) häufig nur in bestimmten Wachstumsphasen oder Entwicklungsstadien der Zelle bzw. des Organismus gebildet werden. Zu betonen ist, dass diese Definition keine Aussagen zur biologischen Funktion von Sekundärmetaboliten macht.

1.1.2 Vorkommen von Sekundärmetaboliten in den Organismenreichen

Die enorme Vielfalt von Verbindungen, die gemäß dem vorherigen Abschnitt als Sekundärmetabolite zu klassifizieren sind, wird daran erkennbar, dass in einer umfassenden Darstellung dieses Themas (Luckner 1990) über 120 verschiedene Sekundärmetabolit-Stoffgruppen unterschieden werden. Diese Verbindungen sind zwar in allen Organismenreichen anzutreffen, kommen jedoch besonders gehäuft in Pflanzen, Pilzen und Bakterien vor. Im Folgenden werden einige ausgewählte Gruppen von Sekundärmetaboliten exemplarisch vorgestellt.

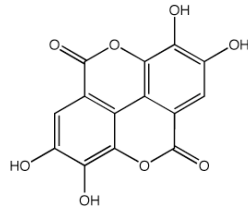
In **Pflanzen** finden sich zahlreiche Sekundärstoffwechselprodukte. Tannine gehören zu den Gerbstoffen und zeichnen sich dadurch aus, dass sie eine ausreichende Zahl von phenolischen Dihydroxy-Gruppen in *ortho*-Stellung besitzen, um biologische Makromoleküle wie Proteine und Pektine quervernetzen zu können (Haslam 1989). Ein Beispiel ist die Ellagsäure (**Abb. 1, 1**), die bei der Hydrolyse von Ellagitanninen entsteht.

Hydroxyzimtsäuren sind Phenylpropanoid-Verbindungen, die sich aus dem Shikimisäure-Stoffwechsel ableiten. Sie besitzen mindestens einen aromatischen C₆-Ring, der eine oder mehrere Hydroxyl-Gruppen trägt (Gross 1981). Es handelt sich also um phenolische Verbindungen. Als Beispiel für diese Stoffgruppe sei hier 4-Cumarinsäure (**Abb. 1, 2**) angeführt. Aus Hydroxyzimtsäure-Derivaten gehen auch die diversen pflanzlichen Lignine durch Polymerisation hervor (Sarkanen und Ludwig 1971).

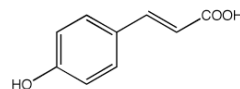
Flavonoide kommen in Pflanzen ubiquitär in überaus großer Strukturvielfalt vor. Sie besitzen eine C₆-C₃-C₆-Grundstruktur, die sich aus zwei aromatischen Ringen A und B und einem Heterozyklus C mit

einem Sauerstoff-Atom aufbaut (**Abb. 1, 3**). Sie werden aus Zwischenprodukten der Shikimisäure-Biosynthese und Malonyl-CoA-Einheiten durch so genannte Chalcon-Synthasen aufgebaut (siehe Abschnitt 1.2.4). Die Grundstruktur unterliegt vielfältigen Modifikationen, woraus sich die verschiedenen Unterklassen der Flavonoide ableiten, zu denen unter anderem die beiden Pflanzenfarbstoff-Gruppen Anthocyane und Flavone gehören (Forkmann und Heller 1999).

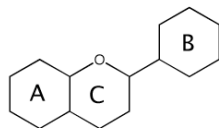
Terpene, synonym auch Isoprene oder Isoprenoide genannt, leiten sich aus Isopren-Einheiten ab, weshalb die Anzahl der C-Atome meist ein Vielfaches von fünf ist.



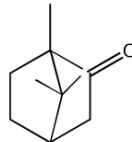
1: Ellagsäure



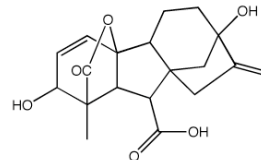
2: 4-Cumarinsäure



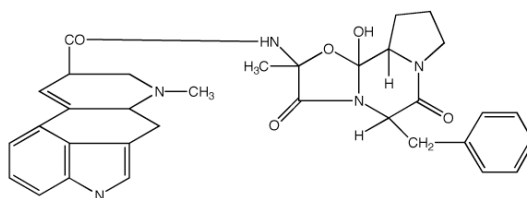
3: Flavonoid-Grundstruktur



4: Campher



5: Gibberellin GA3



6: Ergotamin

Abb. 1: Beispiele für Sekundärmetabolit-Strukturen.

Alle gezeigten Metabolite kommen in Pflanzen vor. Gibberelline werden auch von Pilzen produziert.

Sie werden von Prenyl-Transferasen und Terpen-Cyclasen aufgebaut (Alonso und Croteau 1993). Nach der Zahl der eingebauten Isopren-Einheiten unterscheidet man u. a. zwischen Hemiterpenen, Monoterpenen, Sesquiterpenen, Diterpenen, Triterpenen, Tetraterpenen und Polyterpenen. Typische Vertreter dieser Stoffgruppe sind Campher (ein bicyclisches Monoterpen, **Abb. 1, 4**), die Gibberelline (tetracyclische Diterpene, **Abb. 1, 5**), die Carotinoide (Tetraterpene) und Kautschuk, der ein Polyterpen darstellt.

Alkaloide sind basische Naturstoffe, die ein oder mehrere Stickstoffatome enthalten. Aufgrund ihrer Strukturvielfalt ist eine systematische Einteilung schwierig. Nach dem jeweiligen Alkaloid-Grundgerüst und den entsprechenden biogenetischen Vorstufen unterscheidet man u. a. Indol-Alkaloide (aus Tryptophan), Terpen-Alkaloide (aus Mevalonsäure), Isochinolin-Alkaloide (aus Phenylalanin oder Tyrosin) und Tropan-Alkaloide (aus Ornithin und Acetat).

In **Pilzen** kommen hauptsächlich vier Typen von Sekundärmetaboliten vor (Keller, Turner und Bennett 2005): Polyketide, nichtribosomale Peptide, Terpene und Indol-Alkaloide. Polyketide und die für ihre Synthese zuständigen Enzyme werden in Abschnitt 1.2 behandelt, die nichtribosomalen Peptide und die sie aufbauenden Synthetasen in Abschnitt 1.3. Zu den Terpenen, die in Pilzen produziert werden, gehören z. B. die Aristocholone, außerdem Carotinoide und Gibberelline. Das im Hinblick auf seine Biosynthese bestuntersuchte Indol-Alkaloid ist Ergotamin (**Abb. 1**, 6) aus *Claviceps purpurea*, ein Mutterkorn-Alkaloid.

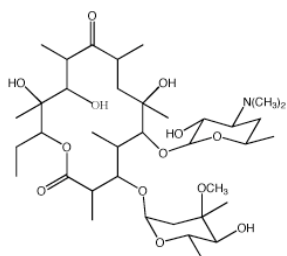
Bakterien synthetisieren ebenfalls eine große Zahl von Sekundärmetaboliten, darunter diverse Terpen-Verbindungen, seltene Zucker-Verbindungen wie das Tetrasaccharid Acarbose, Indole und Glycopeptide (Luckner 1990). Historisch betrachtet, wurde das Sekundärstoffwechsel-Potenzial der Bakterien in seinem Umfang später als das der Pflanzen entdeckt, weil viele dieser Stoffe aus der Bakterienzelle ausgeschieden werden, während Pflanzen sie meist in den Zellen, oft auch in großer Menge, ablagern. Die Erforschung des bakteriellen Sekundärstoffwechsels hat sich in letzter Zeit auf zwei Arten von darin vorkommenden Enzymsystemen konzentriert: zum einen auf die unterschiedlichen Formen von Polyketidsynthasen, zum anderen auf nichtribosomale Peptidsynthetasen, dies gerade auch vor dem Hintergrund, dass viele der von diesen Enzymen hergestellten Verbindungen medizinisch relevante biologische Aktivitäten aufweisen (Bode und Müller 2005). In den folgenden Abschnitten wird näher auf diese beiden Enzymsysteme eingegangen.

1.2 Polyketidsynthasen

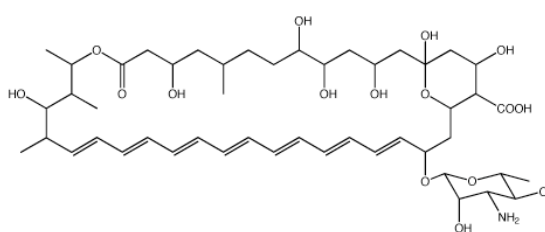
Polyketide ist eine Sammelbezeichnung für Verbindungen, die aufgrund des Biosynthese-Ablaufs zahlreiche Keto-Gruppen enthalten und daher als Polyketon-Verbindungen angesprochen werden können. Das Grundgerüst der Polyketide wird aus Acetyl-CoA- und Malonyl-CoA-Einheiten aufgebaut, wobei die einzelnen Syntheseschritte weitgehende Übereinstimmung mit der aus der Fettsäuresynthese bekannten Reaktionssequenz aufweisen (Hopwood und Sherman 1990). Bei jedem Kettenverlängerungsschritt entsteht eine neue C-C-Bindung unter Ausbildung einer Keto-Gruppe, woraus sich die Bezeichnung Polyketide ableitet. Allerdings schließen sich häufig weitere Reaktionen an, die zu einer Veränderung der Oxidationsstufe am Carbonyl-Kohlenstoff führen, so dass die Polyketo-Struktur bei vielen Vertretern dieser Verbindungsklasse nicht mehr klar zu erkennen ist. Polyketide werden entweder strukturell oder auf Basis ihrer Biosynthese-Wege eingeteilt. In struktureller Hinsicht kann zwischen reduzierten Polyketiden und weitgehend unreduzierten Polyketiden mit aromatischem Charakter unterschieden werden (Weissman und Leadlay 2005). Die reduzierten Polyketide lassen sich weiter in zwei Hauptgruppen unterteilen: makrozyklische Verbindungen, die große Lacton- oder Lactam-Ringe enthalten wie beispielsweise das Makrolid-Antibiotikum Erythromycin A (**Abb. 2**, 1) und die Polyen-Verbindung Amphotericin B (**Abb. 2**, 2)

und zum anderen Polyether-Verbindungen, wie sie z. B. durch das Antibiotikum Monensin A repräsentiert werden (**Abb. 2, 3**). Allerdings gibt es auch viele Polyketide, die sich nur schwer strukturell klassifizieren lassen, da sie unterschiedliche Merkmale in sich vereinigen. Als Beispiel sei der Antitumor-Wirkstoff C-1027 genannt (**Abb. 2, 4**), der zur Gruppe der Eneidyne gehört.

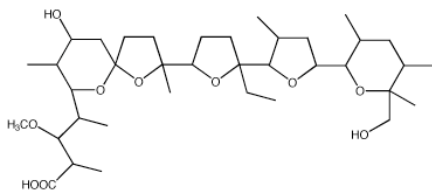
Die ihrer Biosynthese dienenden Enzymkomplexe werden als Polyketidsynthasen (PKS) bezeichnet und in Anlehnung an die Systematik der Fettsäuresynthasen (FAS, von englisch *fatty acid synthase*) in struktureller und funktioneller Hinsicht in drei Hauptklassen, die so genannten Typ I-, Typ II- und Typ III-PKS, eingeteilt. Aufgrund der Ähnlichkeiten zwischen PKS und FAS, die sowohl biochemisch als auch evolutionär bedeutsam sind, wird zunächst ein Überblick über die unterschiedlichen FAS-Systeme gegeben, anschließend werden die drei verschiedenen PKS-Typen vorgestellt.



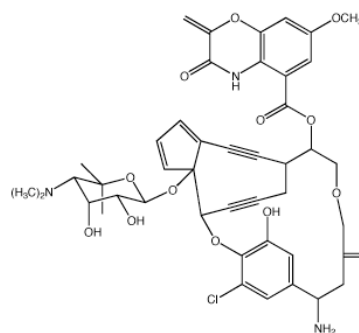
1: Erythromycin A



2: Amphotericin B



3: Monensin A



4: C-1027

Abb. 2: Beispielstrukturen für die unterschiedlichen Polyketid-Klassen.

Die Produzenten der abgebildeten Verbindungen sind: (1) *Saccharopolyspora erythrea*; (2) *Streptomyces nodosus*; (3) *Streptomyces cinnamomensis*; (4) *Streptomyces globisporus*

1.2.1 Fettsäuresynthasen

Fettsäuren treten als Bestandteile von Membranlipiden und Energiespeicher-Substanzen ubiquitär in allen Organismenreichen auf. Dementsprechend finden sich FAS in allen bekannten Organismen. Während die beteiligten enzymatischen Aktivitäten bei allen FAS identisch sind, zeigen sich in Abhängigkeit von der

Organismengruppe Unterschiede in der strukturellen Komposition der Enzyme. Üblicherweise werden die FAS in einen Typ I und einen Typ II unterteilt (Wakil 1961). FAS vom Typ I (FAS-I) sind dadurch charakterisiert, dass sich mehrere enzymatische Aktivitäten auf derselben Polypeptid-Kette befinden, in der sie strukturell und funktionell gegeneinander abgrenzbare Domänen bilden. Dieser Typ findet sich in Tieren, Pilzen und einigen Bakterien-Gattungen (Schweizer und Hofmann 2004). Bei FAS vom Typ II (FAS-II) hingegen liegen die Enzymaktivitäten in Form einzelner, distinkter Proteine vor, die zu einem Multienzymkomplex assoziieren. FAS-II sind die typischen prokaryotischen Systeme zur Fettsäuresynthese; sie kommen in allen Bakterien, auch den Archaea, und in den Chloroplasten der Pflanzen sowie allgemein in den Mitochondrien eukaryotischer Organismen vor.

1.2.1.1 Biochemie der Fettsäuresynthese

Die Biosynthese der Fettsäuren wurde bereits in den 60er Jahren des 20. Jahrhunderts durch Arbeiten mit bakteriellen und Hefe-Systemen in ihren biochemischen Einzelschritten aufgeklärt (Toomey und Wakil 1966; Lynen 1969; Vagelos, Alberts und Majerus 1969). In jüngerer Zeit wurden mehrere umfassende Übersichtsartikel zu diesem Themenkomplex publiziert (unter anderem Rock und Jackowski 2002; Smith, Witkowski und Joshi 2003; Chirala und Wakil 2004). Die einzelnen Reaktionen der Fettsäuresynthese sind schematisch in **Abb. 3** zusammenfassend dargestellt. Die als Bausteine dienenden Acetyl-CoA-Einheiten werden durch die Acetyl-CoA-Carboxylase Biotin-abhängig carboxyliert, wodurch Malonyl-CoA entsteht. Zur Initiation der Biosynthese wird eine Acetyl-Gruppe von Acetyl-CoA auf eine Cystein-Seitengruppe im aktiven Zentrum der Ketoacylsynthase (KS) übertragen und eine Malonyl-Gruppe auf die 4'-Phosphopantetheinyl-Gruppe des Acyl-Träger-Proteins (ACP, von englisch *acyl carrier protein*). (Siehe hierzu auch Abschnitt 1.2.6.) Beide Übertragungen werden von demselben Enzym, der Malonyl-/Acetyl-Transferase (MAT) katalysiert. Im nächsten Schritt übernimmt die KS die Bildung einer neuen C-C-Bindung, die als Claisen-Reaktion abläuft. Dabei tritt als Zwischenstufe ein Enolat-Anion auf, das unter Abspaltung von CO₂ aus der Malonyl-Gruppe gebildet wird. Dass es sich bei der Bindungsbildung um eine Claisen-Reaktion handelt, wird an der stereochemischen Inversion am C-2 sichtbar. Die erste Kettenverlängerung führt also zu einer 3-Ketobutyryl-Einheit, die weiterhin als Thioester kovalent an das ACP gebunden ist. Im nächsten Schritt wird diese C4-Einheit durch die Ketoacylreduktase (KR) unter Beteiligung des Cosubstrats NADPH zum (3*R*)-Hydroxyacyl-ACP (oder synonym: D-β-Hydroxyacyl-ACP) reduziert. Es folgen die durch die Dehydratase (DH) katalysierte Abspaltung von Wasser als *syn*-Eliminierung des 2*S*-Wasserstoffs und der 3*R*-Hydroxyl-Gruppe, wodurch ein *trans*-Enoyl-ACP entsteht, und abschließend die von der Enoylreduktase (ER) durchgeführte Reduktion zum Butyryl-ACP, wobei die Reduktionsäquivalente wiederum einem NADPH entstammen. Zur Vorbereitung des nächsten Kettenverlängerungsschritts wird die Butyryl-Gruppe auf die Cystein-Seitenkette im aktiven Zentrum der KS transferiert und das damit frei gewordene 4'-Phosphopantetheinyl des ACP erneut mit einer Malonyl-Gruppe beladen. Der Reaktionszyklus aus Bindungsbildung durch die KS, Keto-Reduktion durch die KR, Eliminierung von Wasser durch die DH und abschließende Enoyl-Reduktion durch die ER wird mehrfach durchlaufen, jeweils unter Einbau einer neuen C2-Einheit. Auf der Stufe des vollständig gesättigten Acyl-ACP kann die fertig synthetisierte Fettsäure hydrolytisch durch eine Thioesterase (TE) vom ACP abgespalten und freigesetzt werden.

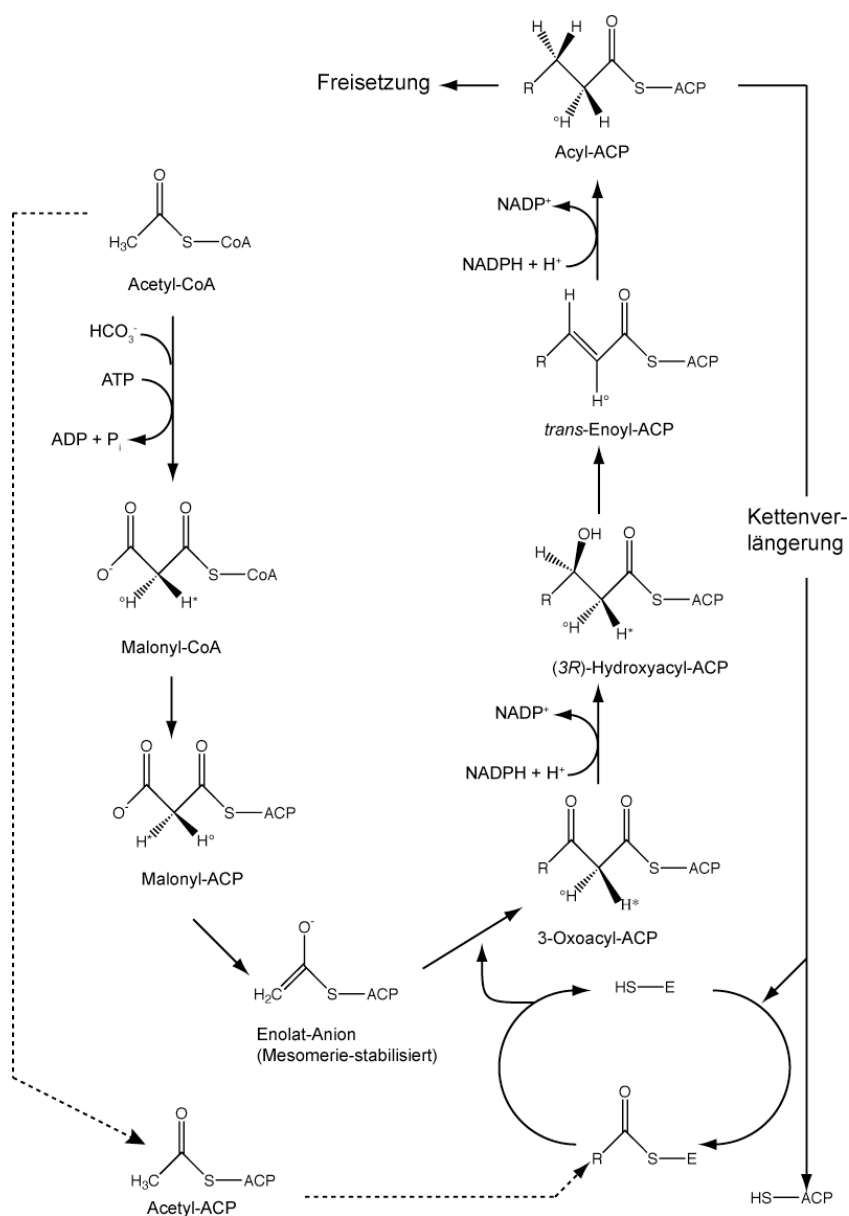


Abb. 3: Schema der Fettsäuresynthese.

Dargestellt ist die Reaktionssequenz der *de novo*-Synthese von Fettsäuren. Die prochiralen H-Atome des Malonyl-CoA sind mit H^o und H^{*} bezeichnet. E steht für den Enzymkomplex, SH für die freie Sulfhydryl-Gruppe des Cystein-Restes im aktiven Zentrum der KS. ACP steht für das Acyl-Träger-Protein. (Nach Metzler 2003, verändert)

1.2.1.2 Struktur der verschiedenen FAS-Typen

Die grundsätzliche Unterscheidung wird zwischen FAS-I und FAS-II getroffen. Wie sich aus **Abb. 4** ergibt, liegen die im vorausgegangenen Abschnitt beschriebenen Enzymaktivitäten bei den FAS-II auf diskreten Polypeptid-Ketten (**Abb. 4, I**).

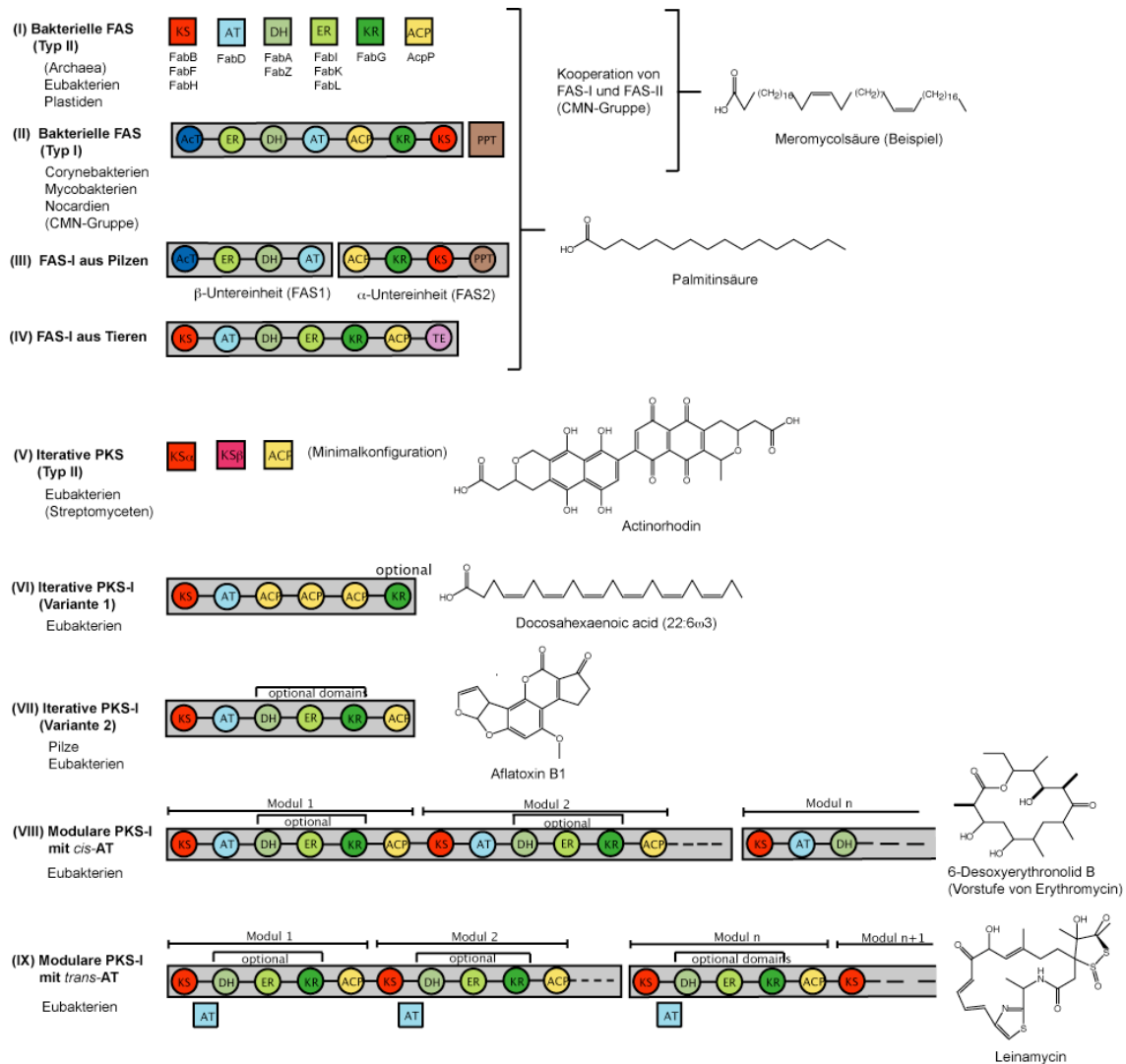


Abb. 4: Vergleichende Übersicht zu den verschiedenen Typen von FAS und PKS.

Polypeptid-Ketten sind als Kästen dargestellt, Domänen innerhalb von Polypeptid-Ketten als Kreise. Optionale Domänen, die nicht immer vorkommen, sind als solche gekennzeichnet. Zu jedem Synthesystem sind Beispielstrukturen von Metaboliten gezeigt. Nicht dargestellt sind modifizierende Enzyme, die zur Synthese des Endproduktes notwendig sind.

Für einige der Komponenten existieren Isoenzym-Formen, die teilweise Unterschiede in der Substratspezifität und in den Enzymaktivitäten aufweisen. So findet man typischerweise in bakteriellen Genomen drei verschiedene KS-Komponenten, FabB, FabF und FabH, die alternativ auch als KS I, KS II bzw. KS III bezeichnet werden (Marrakchi, Zhang und Rock 2002). FabH ist zuständig für die Initiation der Fettsäuresynthese, d.h. es katalysiert die erste Claisen-Reaktion zur Verknüpfung der beiden Starter-Bausteine (Rock und Jackowski 2002). FabH unterscheidet sich von den beiden anderen bakteriellen KS-Enzymen, die ihrer Funktion entsprechend häufig als Elongations-KS bezeichnet werden, dadurch, dass es als Substrat CoA-Thioester anstatt Acyl-ACP benutzt (Jackowski und Rock 1987). FabB und FabF hingegen unterscheiden sich in ihrer Substratspezifität dahingehend, dass sie bevorzugt Acyl-Reste unterschiedlicher Länge umsetzen (Rock und Jackowski 2002). Darüber hinaus existieren auch

spezialisierte Substratspezifitäten zwischen einzelnen Bakteriengattungen. Während in Bakterien, die ausschließlich unverzweigte Fettsäuren produzieren, FabH nur Acetyl- und Propionyl-CoA umsetzen kann (Heath und Rock 1996), zeigen die FabH-Proteine aus Bakterien, die zur Synthese verzweigtkettiger Fettsäuren befähigt sind, hohe Spezifität für verzweigtkettige Acyl-CoA-Varianten, die aus dem Aminosäure-Katabolismus stammen (Han, Lobo und Reynolds 1998). Ein weiteres Beispiel für bakterielle FAS-Isoenzyme mit Unterschieden hinsichtlich ihrer enzymatischen Funktionalität sind die beiden DH-Isoformen FabA und FabZ. Letzteres Enzym katalysiert ausschließlich die Abspaltung eines Wasser-Moleküls, FabA hingegen kann zusätzlich eine Isomerisierung von Doppelbindungen bewerkstelligen, was in Bakterien den ersten Schritt der Biosynthese von ungesättigten Fettsäuren darstellt (Marrakchi, Zhang und Rock 2002).

Im Laufe der Zeit konnten dreidimensionale Strukturdaten von fast allen FAS-II-Komponenten erhalten werden, darunter die KS-Isoformen FabB (Olsen *et al.* 2001), FabF (Huang *et al.* 1998) und FabH (Qiu *et al.* 1999), die AT-Komponente FabD (Serre *et al.* 1995), eines der Enzyme mit ER-Aktivität (Baldock *et al.* 1996), FabA (Leesong *et al.* 1996), die KR-Komponente FabG (Price *et al.* 2001) sowie ACP (Holak *et al.* 1988). Allerdings liegen bislang noch keine Daten dazu vor, welche Arten von Multienzymkomplexen die einzelnen Proteine der bakteriellen Fettsäuresynthese unter physiologischen Bedingungen tatsächlich bilden.

FAS-I kommen in verschiedenen Varianten vor, deren vereinigendes Merkmal ist, dass alle oder zumindest mehrere der Enzymaktivitäten auf derselben Polypeptid-Kette integriert vorliegen. Die Strukturvarianten lassen sich in folgende Untergruppen einordnen: FAS-I aus Tierorganismen (**Abb. 4, IV**), diejenige aus Pilzen (**Abb. 4, III**) und FAS-I bakterieller Herkunft (**Abb. 4, II**). Beim Vergleich der beiden eukaryotischen Varianten fallen bezüglich der Gesamtarchitektur und der Domänenanordnung drei Unterschiede ins Auge. Zum einen sind bei den tierischen FAS-I sämtliche enzymatischen Funktionen auf einem Protein vereint, während FAS-I aus Pilzen eine β -Untereinheit (FAS1) und eine α -Untereinheit (FAS2) aufweisen. Weiterhin enthalten FAS-I aus Tieren auch die TE-Aktivität zur hydrolytischen Abspaltung der Fettsäure aus dem Komplex. Diese ist bei Pilzen nicht Teil des Komplexes, sondern liegt als getrenntes Protein vor. Andererseits enthält die Pilz-FAS zusätzlich eine Phosphopantetheinyl-Aktivität (PPTase) zur Aktivierung des ACP. Phosphopantetheinyl-Transferasen sind Enzyme, die posttranslational den Transfer einer 4'-Phosphopantetheinyl-Gruppe von Coenzym A auf ACP-Domänen von FAS und PKS und ebenso auf die PCP-Domänen von NRPS katalysieren (Gocht und Marahiel 1994; Reuter *et al.* 1999; Mootz, Finking und Marahiel 2001). Schließlich fällt die Anordnung der Domänen innerhalb des Proteins unterschiedlich aus. Tierische FAS-I zeigen die Domänenabfolge KS-AT-DH-ER-KR-ACP-TE, wohingegen man bei Pilzen die Anordnung AcT-ER-DH-AT auf der β -Untereinheit und ACP-KR-KS-PPT auf der α -Untereinheit findet, wobei AcT eine spezialisierte Acetyltransferase-Aktivität bezeichnet, die bei der Initiation der Fettsäuresynthese eine Rolle spielt.

In Bakterien sind FAS-I nicht ubiquitär verbreitet, sondern kommen nur in bestimmten Gattungen innerhalb der Actinobakterien vor, nämlich in Corynebakterien, Mycobakterien und Nocardien, die man als so genannte CMN-Gruppe zusammenfasst (Brennan und Nikaido 1995). Die Vertreter dieser Gattungen haben gemein, dass sie eine im Vergleich mit anderen Bakterien höchst komplexe Zellwandstruktur besitzen, die eine Reihe von Besonderheiten aufweist: (1) Das Peptidoglykan-Gerüst enthält als Diaminosäure *meso*-Diaminopimelinsäure, (2) es tritt das einzigartige Polysaccharid Arabinogalactan auf,

und (3) dieses ist durch charakteristische, sehr langkettige Fettsäuren, die so genannten Mycolsäuren, substituiert (Minnikin 1982). Die Bakterien der CMN-Gruppe besitzen auch das typisch prokaryotische FAS-II-System, sie bilden jedoch insofern eine bemerkenswerte Ausnahme innerhalb der Prokaryoten, als sie zur normalen Fettsäuresynthese nicht auf dieses, sondern auf ihre FAS-I zurückgreifen (Kikuchi, Rainwater und Kolattukudy 1992). Die FAS-II dient statt dessen dazu, die von der FAS-I als Ausgangsmaterial gelieferten Fettsäuren mittlerer Länge in die etwa 40 bis 90 C-Atome umfassenden Mycolsäuren zu überführen (Schweizer und Hofmann 2004). Zieht man diese bakteriellen FAS-I in den weiter oben angestellten Vergleich mit ein, so fällt insbesondere die völlige Übereinstimmung mit den FAS-I aus Pilzen im Hinblick auf die Anordnung der Domänen auf (vgl. **Abb. 4**, II und III).

Bereits seit längerem war bekannt, dass Pilz-FAS-I in der Zelle als $\alpha_6\beta_6$ -Heterododecamer (zusammengefasst in Schweizer und Hofmann 2004) und tierische FAS-I als α_2 -Homodimer (zusammengefasst in Smith, Witkowski und Joshi 2003) vorliegen. Aufgrund der mit Dibromopropanon, einem Reagenz für das „cross linking“ von Thiol-Gruppen, erhaltenen Ergebnissen (Stoops und Wakil 1981) ging man lange Zeit davon aus, dass die beiden Monomere im tierischen Komplex in einer so genannten „Kopf-an-Schwanz“-Anordnung vorliegen, also sich dergestalt zusammenlagern, dass der Amino-Terminus des einen Proteins in direktem Kontakt zum Carboxy-Terminus des anderen steht, wobei die beiden Proteine quasi ausgestreckt nebeneinander liegen. Die Untersuchung der Komplementation in Mutanten (Rangan, Joshi und Smith 2001) und von FAS-Komplexen mit eingeschränkter Funktionalität (Joshi *et al.* 2003) sowie verbesserte „cross linking“-Studien (Witkowski *et al.* 1999) führten allerdings zu einer Revision des ursprünglichen Modells. Es wurde eine Anordnung vorgeschlagen, bei der die KS- und AT-Domänen näher am Zentrum des Komplexes liegen und somit Zugang zu den ACP-Domänen beider Polypeptid-Ketten haben.

Vor kurzem gelang es, beide FAS-I-Subtypen zu kristallisieren und Strukturmodelle mit einer Auflösung von jeweils etwa 5 Angström zu erhalten (Jenni *et al.* 2006; Maier, Jenni und Ban 2006). Die Kristallstruktur der tierischen FAS-I bestätigte weitgehend die neueren Vorschläge.

1.2.2 Polyketidsynthesen des Typs I

In Bezugnahme auf die bei FAS übliche Terminologie werden solche PKS als PKS des Typs I (PKS-I) bezeichnet, bei denen die enzymatischen Aktivitäten auf einer Polypeptidkette integriert vorliegen. Diese PKS-I können wiederum in zwei Subtypen eingeordnet werden, wobei sowohl strukturelle wie auch funktionelle Gesichtspunkte zur Einteilung herangezogen werden (Staunton und Weissman 2001). Zum einen gibt es modular aufgebaute Systeme, die dadurch ausgezeichnet sind, dass jeweils ein so genanntes Modul für den Einbau eines Monomeren in die wachsende Polyketid-Kette verantwortlich ist und die dafür erforderlichen Enzymaktivitäten in sich vereint. Mehrere Module werden im Syntheseprozess hintereinander geschaltet, woraus sich, bildlich gesprochen, ein Fließband-Ablauf ergibt. Die andere Gruppe von PKS-I besteht dagegen ähnlich den FAS-I nur aus einem vollständigen Satz von Enzymdomänen, die wiederholt Monomere in das wachsende Polyketid einbauen, weshalb sie auch als iterative PKS-I bezeichnet werden.

1.2.2.1 Modulare PKS-I

1.2.2.1.1 Modularität als Basis chemischer Variabilität

Die Module der modularen PKS-I sind funktionelle Einheiten, die jeweils für den Einbau eines Monomeren zuständig sind. Die Gesamtstruktur stellt sich als Multienzym-Komplex dar, bestehend aus der Gesamtheit aller Module in linearer Abfolge (siehe **Abb. 4**, VIII). Dabei können die Module entweder als distinktes Protein organisiert sein oder in Form von Proteinen, die mehrere Module enthalten. Wie aus **Abb. 4** ersichtlich ist, verfügen PKS-I über exakt die gleichen enzymatischen Aktivitäten wie FAS. Im Unterschied zu diesen sind jedoch nicht alle diese Aktivitäten in jedem Modul vorhanden. Nur die ACP-Domäne als Acyl-Gruppen-Träger, die AT-Domäne zu ihrer Beladung und die KS-Domäne zur C-C-Bindungsbildung sind obligate Bestandteile, während die Domänen, die dem reduktiven Umbau des Polyketid-Gerüsts bis zur völligen Sättigung dienen, optionale Komponenten darstellen. Daher finden sich in modularen PKS-I vier verschiedene Modultypen, im Folgenden mit A, B, C und D bezeichnet.

Die Minimalkonfiguration besteht aus der Anordnung KS-AT-ACP und kann durch das Hinzutreten weiterer Domänen ergänzt werden, was entsprechende Auswirkungen auf die Struktur der Polyketid-Kette hat (siehe **Abb. 5**). Das Wirken eines Minimalmoduls (Typ A) führt zu einer Keto-Gruppe in β -Stellung, der Modultyp B zu einer Hydroxyl-Gruppe in der β -Position, der Modultyp C ergibt eine Methenyl-Gruppe, der Modultyp D schließlich die komplett reduzierte Methylen-Gruppe. Diese Spezifizierung von Modultypen gilt jedoch nur für Elongationsmodule, d.h. solche Module, die nach dem Synthesestart die Polyketid-Kette verlängern. Die zu Beginn der Synthese beteiligten so genannten Beladungsmodule (englisch: *loading modules*) sind auf andere Weise zu klassifizieren (siehe Abschnitt 1.2.2.1.2).

Es existieren AT-Domänen mit unterschiedlicher Substratspezifität. Der eine Typ von AT-Domänen benutzt das aus der Fettsäuresynthese bekannte Malonyl-CoA, während der zweite Typ Methylmalonyl-CoA als Substrat verwendet. Letzteres wird aus Propionyl-CoA gebildet, das beim Abbau von Fettsäuren mit einer ungeraden Zahl von C-Atomen und im Katabolismus der Aminosäuren Methionin, Valin und Isoleucin entsteht. Propionyl-CoA wird unter der Katalyse der Biotin-abhängigen Methylmalonyl-CoA-Carboxylase (EC 4.1.1.41) unter ATP-Verbrauch zum D-Methylmalonyl-CoA (synonyme Bezeichnung: (2R)-Methylmalonyl-CoA) carboxyliert und danach mit Hilfe der Methylmalonyl-CoA-Epimerase (EC 5.1.99.1) zum L-Methylmalonyl-CoA (synonyme Bezeichnung: (2R)-Methylmalonyl-CoA) isomerisiert, welches wiederum im Primärstoffwechsel mittels der Methylmalonyl-CoA-Mutase (EC 5.4.99.2) in Succinyl-CoA überführt wird. Alle bisherigen Untersuchungen deuten darauf hin, dass PKS-I ausschließlich das L-Isomer als Substrat heranziehen (Weissman *et al.* 1997). Die Verwendung von Methylmalonyl-CoA bei der Synthese von Polyketiden veranschaulicht die allgemeine Beobachtung, dass sich Sekundärmetabolite von Verbindungen des Primärstoffwechsels ableiten (siehe Abschnitt 1.1.1).

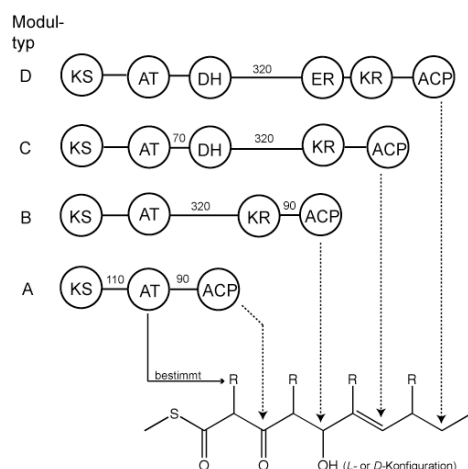


Abb. 5: Die unterschiedlichen Modultypen in modularen PKS-I.

Domänen sind als Kreise dargestellt, die verbindenden Interdomänen-Bereiche als Striche. Die Zahlen geben typische Längen der Interdomänen-Bereiche bezogen auf die Aminosäurereste in der Polyketid-Kette an. Jeder Modultyp führt zu einer anderen Teilstruktur innerhalb des Polyketid-Grundgerüsts. R bezeichnet allgemein ein Wasserstoff-Atom oder einen Alkyl-Rest. Abkürzungen: KS, Ketoacylsynthase; AT, Acyltransferase; DH, Dehydratase; KR, Ketoreduktase; ACP, Acyl-Träger-Protein.

Eine systematische Analyse auf der Grundlage von 187 AT-Domänen aus charakterisierten PKS-I ergab, dass die Substratspezifität im Wesentlichen von 13 Aminosäure-Resten im aktiven Zentrum und weiteren 11 hoch konservierten Aminosäure-Positionen außerhalb des aktiven Zentrums bestimmt wird (Yadav, Gokhale und Mohanty 2003). AT-Domänen, die Malonyl-CoA verwerten, zeigen das charakteristische Aminosäure-Motiv QQGHS[LVIFAM]GR[FP]H [ANTGEDS][NHQ]V im aktiven Zentrum, während man bei solchen AT-Domänen, die spezifisch für Methylmalonyl-CoA sind, das Motiv QQGHS[QMI]GRSHT[NS]V findet. Hingegen war eine klare Erfassung der Spezifität für seltenere Substrate wie etwa Ethylmalonat und 2-Methylbutyrat, die hauptsächlich in Beladungsmodulen, daneben aber auch bisweilen in Elongationsmodulen vorkommen, aufgrund der zu kleinen Datenbasis nicht möglich.

Auch KR-Domänen treten in unterschiedlichen Varianten auf. Anhand von Struktur-aufgeklärten Polyketid-Verbindungen wird klar, dass die Hydroxyl-Gruppen, die unter Mitwirkung der KR-Domänen aus der ursprünglichen Keto-Gruppe hervorgehen, beide möglichen stereochemischen Konfigurationen aufweisen können (Übersicht bei Rawlings 2001). Eine Gegenüberstellung der entsprechenden KR-Domänen in einem Alignment der Aminosäure-Sequenzen erlaubte deren Aufteilung in zwei Gruppen aufgrund einiger Positionen mit ausreichend konsistenten Unterschieden (Caffrey 2003). Das Auftreten einer D-Hydroxyl-Gruppe korreliert mit der Existenz eines konservierten Aspartat-Rests als Teil eines Motivs mit der Abfolge Leu-Asp-Asp (LDD-Motiv), während der andere Typ von KR-Domänen einen charakteristischen Tryptophan-Rest aufweist und zur Bildung einer L-Hydroxyl-Gruppe führt. Dieser Zusammenhang ist in **Abb. 5** zusammengefasst. Dass DH-Domänen nur bei Vorliegen einer D-Hydroxyl-

Gruppe Wasser abspalten können, scheint der Grund zu sein, weshalb in Modulen vom C-Typ normalerweise nur KR-Domänen, die das LDD-Motiv enthalten, gefunden werden (Wu *et al.* 2005).

Zusammenfassend ergeben sich also folgende Unterschiede zwischen FAS und PKS-I: (1) Während in FAS nach Einbau einer neuen C2-Einheit immer die gesamte Reaktionsfolge bis zum Erreichen einer vollständig gesättigten Kohlenwasserstoff-Kette durchlaufen wird, können in PKS-I durch Weglassen bestimmter Schritte variable Ketten gebildet werden. (2) PKS-I können beide stereochemischen Konfigurationen beim Einführen einer Hydroxyl-Gruppe in β -Position generieren. (3) Aufgrund der Verwertbarkeit anderer Substrate als Malonyl-CoA durch AT-Domänen können in der α -Position verschiedene Alkyl-Ketten vorkommen. Damit besitzen PKS-I das Potential, eine überaus große Zahl unterschiedlicher Polyketid-Grundgerüste aufzubauen. Eine PKS-I mit sechs Elongationsmodulen kann theoretisch etwa 100.000 verschiedene Strukturen erzeugen (Gonzalez-Lergier, Broadbelt und Hatzimanikatis 2005).

Das Zusammenwirken der Module in der Art eines Fließband-Prozesses lässt sich anhand des Beispiels der Erythromycin-PKS aus dem Actinobakterium *Saccharopolyspora erythraea* verdeutlichen, die als erste in den einzelnen biochemischen Schritten aufgeklärt wurde (**Abb. 6**). Dieser Multienzym-Komplex bildet die Verbindung 6-Desoxyerythronolid B, die das Aglycon-Zwischenprodukt der Erythromycin-Synthese darstellt. Die Gene dieser Synthase wurden Anfang der 90er Jahre des 20. Jahrhunderts unabhängig voneinander durch zwei Forschergruppen identifiziert (Cortes *et al.* 1990; Tuan *et al.* 1990) und in der Folgezeit ausführlich untersucht (Donadio *et al.* 1991; Bevit *et al.* 1992).

Aufgrund des Synthesemechanismus ergibt sich die Struktur des Polyketid-Rückgrats unmittelbar aus der Abfolge und der Domänen-Zusammensetzung der Module innerhalb des PKS-Komplexes. Dieser Zusammenhang wird als Colinearitätsregel bezeichnet (Staunton und Weissman 2001).

1.2.2.1.2 Beladungsmodule

Der erste Baustein einer Biosynthese an PKS-I wird auf so genannte Beladungsmodule übertragen. Diese unterscheiden sich von den im vorherigen Abschnitt beschriebenen Elongationsmodulen in zweifacher Hinsicht: Zum einen weichen sie in ihrem Domänen-Aufbau ab bzw. besitzen besondere Domänen-Varianten. Darüber hinaus zeigen sie eine viel größere Variabilität an möglichen Substraten. Die eine Form von Beladungsmodulen enthält keine KS-Domäne, sondern besteht nur aus einer AT und einer ACP-Domäne und wurde zunächst für die Standardform gehalten (Staunton und Wilkinson 1998). Mit der zunehmenden Zahl an PKS-I-Sequenzen stellte sich jedoch heraus, dass es eine weitere Art von Beladungsmodulen gibt. Diese ähnelt einem Minimal-Modul der Konfiguration KS-AT-ACP, enthält allerdings einen besonderen Typ von KS-Domänen, bei denen ein essentielles Serin im aktiven Zentrum durch Glutamin ersetzt ist (Kao *et al.* 1996), weshalb auch von KSQ-Domänen gesprochen wird. Diese speziellen KS-Domänen können zwar keine Claisen-Reaktion zwischen zwei Carbonyl-Bausteinen katalysieren, zeigen jedoch eine Decarboxylase-Aktivität (Bisang *et al.* 1999). Man geht heute davon aus, dass die kürzeren AT-ACP-Beladungsmodulen nicht-carboxyliertes Acyl-CoA als Substrat verwenden (Wiesmann *et al.* 1995), während diejenigen Beladungsmodulen, die den Aufbau KSQ-AT-ACP haben,

carboxylierte Acyl-CoA-Formen benutzen und diese dann mittels der KSQ-Aktivität decarboxylieren (Sun *et al.* 2003).

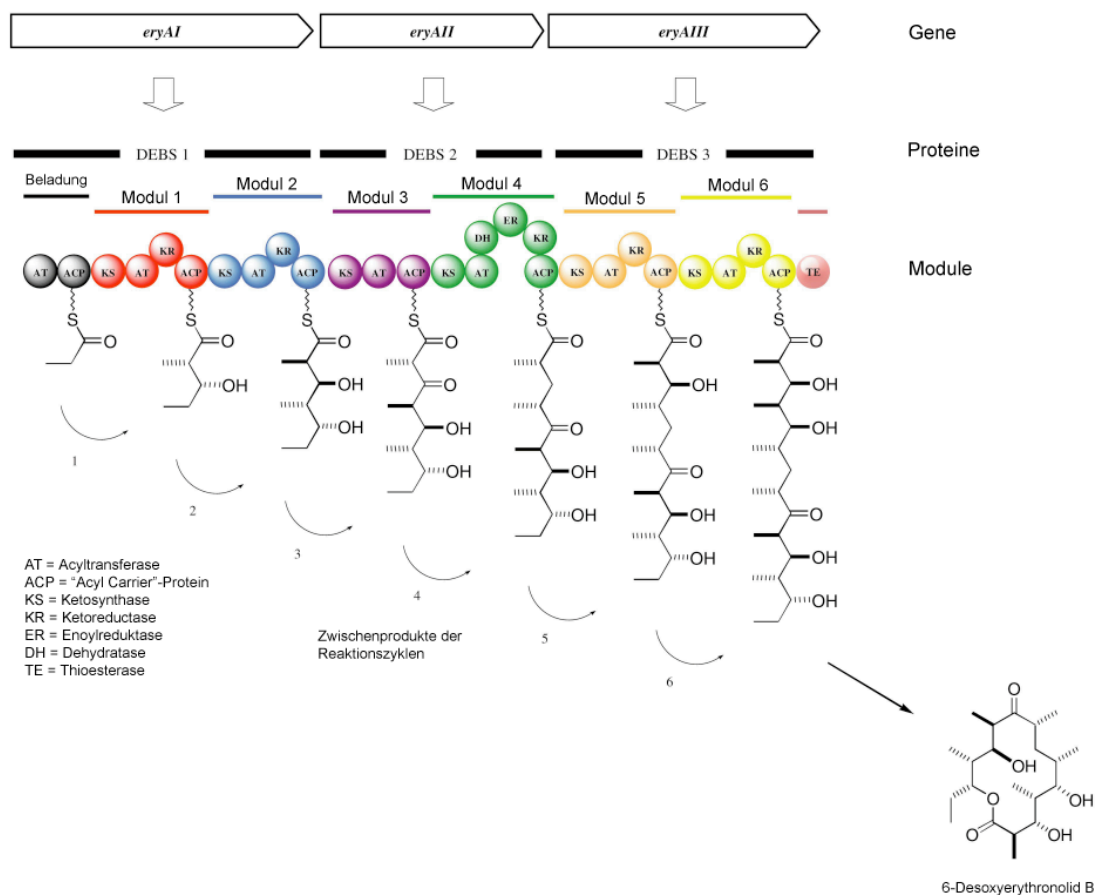


Abb. 6: Biosynthese von 6-Desoxyerythronolide B.

Proteine sind durch schwarze Balken dargestellt. DEBS1 besitzt drei Module, von denen das erste ein Beladungsmodul ist, DEBS2 und DEBS 3 umfassen je zwei Elongationsmodule. Die Module sind farblich gegeneinander abgehoben. Die Kreise stellen die funktionellen Domänen der Module dar. (Nach Staunton und Weissman 2001, verändert)

Die AT-Domänen von Beladungsmodulen sind bezüglich der Substratspezifität weitaus vielfältiger als diejenigen der Elongationsmodule (Moore und Hertweck 2002). Neben den dort ganz überwiegend verwendeten Substraten Malonyl-CoA und Methylmalonyl-CoA, sind eine Reihe spezieller Verbindungen beschrieben worden, die als Starter-Bausteine der Polyketidsynthese dienen können. Eine Auswahl ist in **Tab. 1** aufgelistet.

Tab. 1: Besondere Substratspezifitäten von Beladungsmodulen ausgewählter PKS-I

Substrat der Beladungsdomäne	Sekundärmetabolit	Literaturreferenz
Isobutyryl-CoA	Avermectin	(Ikeda und Omura 1997)
2-Methylbutyryl-CoA	Avermectin	(Ikeda und Omura 1997)
3-Amino-5-Hydroxybenzoyl-CoA	Rifamycin	(August <i>et al.</i> 1998)
3,4-Dihydroxycyclohexylcarbonyl-CoA	Rapamycin	(Schwecke <i>et al.</i> 1995)
p-Nitrobenzoat	Aureothin	(He und Hertweck 2003)
Benzoyl-CoA	Soraphen	(Wilkinson <i>et al.</i> 2001)
3-Methylbutyryl-CoA	Myxothiazol	(Silakowski <i>et al.</i> 1999)

1.2.2.1.3 Modulare PKS-I mit *trans*-AT-Aktivität

Die Modultypen der weit überwiegenden Mehrheit bislang bekannter PKS-I besitzen eine AT-Domäne (**Abb. 4**, IX). Vor einigen Jahren wurde erstmals eine modulare PKS-I beschrieben, deren Module zwar im Prinzip dem üblichen Schema entsprechen, die jedoch keine AT-Domänen aufweisen (Piel 2002). Sie konnte einem bakteriellen Symbionten des Käfers *Paederus fuscipes* zugeordnet werden und synthetisiert das Antitumor-Polyketid Pederin. Es wurde vorgeschlagen, dass bei fehlenden integrierten AT-Domänen in *cis*-Stellung die Module wahrscheinlich durch eigenständige Proteine mit AT-Aktivität beladen werden (Piel 2002). In der Folgezeit konnten weitere Vertreter solcher modularen PKS-I gefunden und teilweise charakterisiert werden: die PKS-I der Biosynthesen von Leinamycin in *Streptomyces globisporus* (Cheng, Tang und Shen 2003), von Mupirocin in *Pseudomonas fluorescens* (El-Sayed *et al.* 2003) und von Lankacidin in *Streptomyces rochei* (Mochizuki *et al.* 2003). Der Mechanismus der Bereitstellung von Synthese-Bausteinen durch *trans*-ständige AT konnte im Fall der Leinamycin-PKS nachgewiesen werden (Cheng, Tang und Shen 2003).

1.2.2.1.4 Abweichungen vom Grundschema der modularen PKS-I-Synthese

Die anhand der als Paradigma dienenden Biosynthese von Erythromycin aufgestellte Colinearitätsregel (Abschnitt 1.2.2.1.1) wird von einigen modularen PKS-I auf unterschiedliche Weise durchbrochen. Zusammenfassungen solcher in Bezug auf das Synthese-Grundschema ungewöhnlichen Funktionsmodi werden in den Übersichtsartikeln von (Wenzel und Müller 2005) und (Moss, Martin und Wilkinson 2004) gegeben. Man findet einerseits Verlängerung von Polyketid-Ketten durch wiederholten Einsatz, d.h. durch den iterativen Gebrauch desselben Moduls, andererseits den gegenteiligen Effekt, dass nämlich durch das Auslassen (englisch: *skipping*) eines Moduls eine kürzere Struktur entsteht, als der Aufbau des Synthese-Systems erwarten lässt. In beiden Fällen kann zwischen eher zufälligen Ereignissen, die zu einem geringen Anteil an Nebenprodukt führen, und einem auf diese Weise festgelegten, also quasi programmierten Modus unterschieden werden.

Ein Beispiel für den zufälligen mehrfachen Moduleinsatz, in der englischsprachigen Literatur häufig als „Stottern“ (englisch: *stuttering*) bezeichnet, liefert die Erythromycin-PKS selbst. Es konnte gezeigt

werden, dass in einem *S. erythraea*-Stamm als Nebenprodukt der Biosynthese des Heptaketids auch Oktaketide gebildet werden, die sich auf eine zweifache Benutzung von Modul 4 zurückführen lassen (Wilkinson *et al.* 2000). Im Gegensatz hierzu wird eines der Module der Biosynthese von Stigmatellin in dem Myxobakterium *Stigmatella aurantiaca* grundsätzlich zweimal iterativ benutzt (Gaitatzis *et al.* 2002). Ebenso ist der iterative Gebrauch des zweiten Moduls bei der Biosynthese von Aureothin in *Streptomyces thioluteus* Bestandteil des regulären Synthese-Ablaufs dieser PKS-I (He und Hertweck 2003). Das unregelmäßige Auslassen eines Moduls und das damit verbundene Auftreten von Nebenprodukten mit kleineren Ringsystemen wurde unter anderem bei der Biosynthese von Epothilonen beobachtet (Hardt *et al.* 2001). Das programmierte Auslassen eines kompletten Moduls wurde am Beispiel der Pikromycin-Synthese aus *Streptomyces venezuelae* intensiv untersucht, wo es in Abhängigkeit von den Kulturbedingungen auftritt (Beck *et al.* 2002). Die Mechanismen der festgelegten Iterativität sind bislang ebenso wenig bekannt wie die Steuerung der programmierten Modul-Auslassung.

1.2.2.2 Iterative PKS-I

Iterativ arbeitende PKS-I sind dadurch charakterisiert, dass sie die Typ-I-spezifische Integration mehrerer funktioneller Domänen auf einer Polypeptid-Kette zeigen, jedoch gewissermaßen nur aus einem einzigen Modul aufgebaut sind (**Abb. 4**, VII). Die lineare Abfolge der Domänen entspricht der Anordnung in modularen PKS-I. Im Gegensatz zu diesen wird allerdings das „Einzelmodul“ wiederholte Male für aufeinander folgende Polymerisationsschritte eingesetzt, was an die mechanistischen Abläufe bei FAS-I erinnert. Der Unterschied zu den FAS-I besteht darin, dass die Elongationszyklen nicht gleichartig ablaufen, sondern in verschiedenen Zyklen unterschiedlicher Gebrauch der enzymatischen Aktivitäten gemacht wird. Eine sehr gut untersuchte iterative PKS-I ist die 6-Methylsalicylsäure-Synthase aus dem Ascomycet *Penicillium patulum*. Auf der Polypeptid-Kette befinden sich alle Domänen-Typen außer einer ER in der Anordnung KS-AT-DH-KR-ACP (Shoolingin-Jordan und Campuzano 1999). Das Produkt 6-Methylsalicylsäure wird aus einer Acetat- und drei Malonat-Einheit aufgebaut, wobei sich an die erste Claisen-Reaktion keine reduktiven Modifizierungen anschließen, sondern sofort der erste Kettenverlängerungsschritt erfolgt. Dieser zweite Zyklus beinhaltet eine Keto-Reduktion und die Elimination eines Wasser-Moleküls. Der dritte Zyklus hingegen verläuft wiederum ohne Reduktionsschritt, woran sich eine Zyklisierung der Polyketid-Kette unter Abspaltung des Produktes anschließt (Shoolingin-Jordan und Campuzano 1999). Es ist nicht bekannt, welcher molekulare Mechanismus die Synthase befähigt, die KR- und die DH-Aktivität ausschließlich im zweiten Zyklus zu benutzen.

Iterative PKS-I wurden zuerst in Pilzen entdeckt, und lange Zeit ging man davon aus, dass diese Enzyme nur dort vorkommen. Die Biosynthesen der Aflatoxine und von Lovastatin sind weitere bekannte Beispiele für gut untersuchte iterative PKS-I aus Pilzen (Staunton und Weissman 2001). Vor einigen Jahren wurde allerdings mit der Avilamycin-Synthase aus *Streptomyces viridochromogenes* erstmals eine PKS dieses Typs beschrieben, die bakteriellen Ursprungs ist (Gaisser *et al.* 1997). Inzwischen sind weitere iterative PKS-I aus Prokaryoten identifiziert worden, unter anderem die 2-Hydroxy-5-methyl-1-naphthonsäure-Synthase aus *Streptomyces neocarzinostaticus*, die eine Komponente des Antitumor-Wirkstoffs Neocarzinostatin produziert (Sthapit *et al.* 2004; Liu *et al.* 2005).

1.2.3 Polyketidsynthesen des Typs II

PKS des Typs II (PKS-II) wurden bislang ausschließlich in Streptomyceten, einer Unterordnung der Actinobakterien gefunden. Sie bestehen aus distinkten Proteinen, die jeweils eine enzymatische Aktivität bereitstellen (Staunton und Weissman 2001) (**Abb. 4**, V). Die Minimalkonfiguration zum Aufbau der Polyketid-Kette, die in allen bekannten PKS-II gefunden wird, umfasst zwei Proteine mit KS-Aktivität, KS_{α} und KS_{β} genannt, sowie ein ACP. Hinzu treten KR, Cyclasen und Aromatasen, die Modifizierungen nach der Fertigstellung des Kohlenstoff-Grundgerüsts einführen. Das am besten untersuchte Beispiel für diesen Synthase-Typ ist die Actinorhodin-PKS aus *Streptomyces coelicolor*. Ebenso wie die Produkte anderer PKS-II handelt es sich um eine aromatische Verbindung mit phenolischem Charakter. Die Biosynthese umfasst folgende Schritte (Richardson und Khosla 1999): Zunächst wird eine nicht-reduzierte Polyketid-Kette aus einer Acetyl- und sieben Malonyl-Einheiten aufgebaut, gefolgt von einer ersten Zyklisierung, zwei Keto-Reduktionen und einer weiteren Zyklisierung. Einige weitere Umsetzungen führen schließlich zum fertigen Endprodukt. Die beiden KS-Komponenten weisen hohe Aminosäure-Sequenzähnlichkeit auf (Siggaard-Andersen 1993), unterscheiden sich jedoch darin, dass KS_{β} den essentiellen Cystein-Rest im aktiven Zentrum hat, während dieser in KS_{α} durch ein Glutamin ersetzt ist (Bisang *et al.* 1999). KS_{β} katalysiert die Claisen-Reaktion zur Kettenverlängerung, die Funktion von KS_{α} blieb hingegen längere Zeit unklar. Ursprünglich vermutete man darin diejenige Komponente, welche die Kettenlänge des Polyketids bestimmt und führte deshalb die Bezeichnung *chain length factor* (CLF) ein, ohne einen konkreten Mechanismus angeben zu können (McDaniel *et al.* 1993). Später stellte sich aber heraus, dass KS_{α} eine Decarboxylase-Aktivität bezüglich Malonyl-ACP besitzt (Bisang *et al.* 1999). Daraus wurde der folgende, mittlerweile allgemein anerkannte Mechanismus abgeleitet: Die KS_{α} -Komponente dient dazu, die nach Beladung des ACP zunächst vorliegenden Malonyl-Gruppe mittels Decarboxylierung in eine Acetyl-Gruppe umzuwandeln, was Voraussetzung für den Start der Polyketidsynthese ist, woraus die neue Bezeichnung Ketteninitiationsfaktor (englische: *chain initiation factor*) abgeleitet wurde (Bisang *et al.* 1999). Im Rahmen derselben Arbeiten wurde auch die Funktion der KSQ-Domänen in den Beladungsmodulen modularer PKS-I untersucht (siehe Abschnitt 1.2.2.1.2).

Das Fehlen eines Enzyms mit AT-Aktivität im Actinorhodin-Cluster wurde zuerst damit erklärt, dass die AT-Komponente der FAS in *S. coelicolor* diese Funktion übernehmen könnte (Revill, Bibb und Hopwood 1995; Summers *et al.* 1995). Wie man inzwischen aufgrund von unabhängig durch mehrere Arbeitsgruppen durchgeführte *in vitro*-Experimenten weiß, besitzt die ACP-Komponente eine AT-Aktivität, für die keine zusätzlichen Proteine notwendig sind (Hitchman *et al.* 1998; Matharu *et al.* 1998; Zhou, Florova und Reynolds 1999). Insofern scheint die Beteiligung von FAS-Komponenten an der PKS-II-Synthese als Erklärung nicht mehr notwendig zu sein, allerdings stehen abschließende Nachweise zu der Frage, wie die Beladung mit Malonyl-Gruppen tatsächlich abläuft, immer noch aus.

Trotz vieler mechanistischer Details der Polyketid-Synthese durch PKS-II, die in den vergangenen 15 Jahren aufgeklärt werden konnten, ist bis heute unklar, wie diese Enzym-Systeme die zwei aus chemischer Sicht schwierigsten Aufgaben auf programmierte Weise bewältigen können, nämlich die exakte Determination der Kettenlänge, also der Zahl der Kettenverlängerungszyklen, und außerdem die

genaue Positionsfestlegung der innermolekularen Rückfaltung zur Bildung des ersten aromatischen Rings (Staunton und Weissman 2001).

1.2.4 Polyketidsynthesen des Typs III

Die Benennung von Polyketidsynthesen als Typ III (PKS-III) nimmt keinen direkten Bezug auf die Einteilung von FAS, denn bei diesen Systemen existiert kein Typ III. Es handelt sich vielmehr um eine Erweiterung der ursprünglich üblichen Bezeichnungsweise, um zu verdeutlichen, dass es neben PKS-I und PKS-II eine weitere Form von Polyketidsynthesen gibt, die sich enzymologisch klar von den beiden erstgenannten abgrenzen lässt. Eine allgemeine Erörterung der terminologischen Schwierigkeiten bei der Klassifizierung von PKS wird in Abschnitt 1.2.5 gegeben. Aufgrund ihres ubiquitären Vorkommens in Pflanzen wurden PKS-III über lange Zeit als typische Pflanzen-PKS aufgefasst und entsprechend benannt. Da jedoch mittlerweile eine Reihe von PKS-III aus Bakterien (Moore und Hopke 2001) und Pilzen (Seshime *et al.* 2005) beschrieben wurde, ist diese Bezeichnung nicht mehr aktuell. Stattdessen wird diese Enzymgruppe unter Bezugnahme auf die Hauptprodukte, die von ihren Vertretern hergestellt wird, die so genannten Chalcone, seit längerem als Chalcon-Synthesen (CHS) bezeichnet. Die Benennung als PKS-III wurde vorgeschlagen, um sie auch begrifflich den Polyketidsynthesen zuzuordnen (Shen und Hutchinson 1993).

PKS-III liegen in der Zelle als homodimerer Komplex vor, die molekulare Masse des Monomeren beträgt etwa 42 bis 45 kDa (Tropf *et al.* 1995). 1999 gelang die Aufklärung der Struktur der CHS aus dem Kleegetränk Alfalfa (*Medicago sativa*, deutsche Bezeichnung: Luzerne) durch Röntgenstruktur-Analyse (Ferrer *et al.* 1999). Das Strukturmodell ergab, dass PKS-III über zwei unabhängige aktive Zentren verfügen. Allgemein besteht die von ihnen katalysierte Gesamtreaktion aus einer Reihe PKS-typischer Kettenverlängerungsreaktionen, wobei Malonyl-CoA-Einheiten als Ausgangsmaterial dienen, und mehreren Zyklisierungsschritten. Die Reaktionsabfolge lässt sich folgendermaßen beschreiben (Jez *et al.* 2001): In der zu Beginn erfolgenden Beladungsreaktion wird die Starter-Einheit von Coenzym A auf das Enzym übertragen, so dass sie als Thioester an eine Cystein-Seitenkette gebunden vorliegt. Nach drei Verlängerungsreaktionen, bei denen jeweils Malonyl-CoA als Substrat auftritt und die über Decarboxylierungen und sich anschließende Claisen-Reaktionen ablaufen, liegt ein Tetraketid-Zwischenprodukt vor. Dieses wird mittels einer intramolekularen Claisen-Reaktion zyklisiert unter Bildung eines Trions, das abschließend zum Chalcon aromatisiert wird. Die Diversität der Produkte beruht darauf, dass verschiedene Verbindungen als Starter-Moleküle dienen, die Kettenlänge über die Zahl der Elongationsschritte variiert werden kann und durch stereospezifische Kontrolle der Zyklisierungs-Reaktionen. PKS-III zeigen bei der Auswahl der Starter-Moleküle große Variabilität, wobei auffällt, dass sich die meisten Verbindungen aus dem Phenylpropanoid-Stoffwechsel ableiten. Neben *para*-Cumarinsäure (4-Cumarinssäure, siehe **Abb. 1, 2**) werden auch die Struktur-verwandten Verbindungen Zimtsäure, Dihydro-Zimtsäure und Caffein sowie Acetat, Isovalerat und *N*-Methylantranilat, jeweils gebunden an Coenzym A, als Substrate benutzt (Schröder 1999). Die Zahl der Elongationsschritte variiert zwischen einem und vier, beträgt jedoch in der überwiegenden Mehrzahl der untersuchten Enzyme drei (Austin und Noel 2003).

Eine ebenfalls sehr gut untersuchte Untergruppe innerhalb der PKS-III sind die aus Pflanzen isolierten Stilben-Synthasen (STS). Sie bilden das gleiche lineare Tetraketid-Zwischenprodukt wie die CHS, führen jedoch eine andere Form der Cyclisierung durch. Während CHS eine intramolekulare Claisen-Reaktion zwischen C6 und C1 (Nummerierung ausgehend von der Thioester-Bindung mit der Cystein-Seitenkette) durchführen, läuft bei den STS eine intramolekulare Aldol-Reaktion zwischen C2 und C7 ab (Tropf *et al.* 1995). Dementsprechend liefern CHS und STS bei Benutzung der gleichen Starter-Einheit trotz der völlig identischen ersten Reaktionsschritte jeweils andere Produkte (Schröder 1999).

1.2.5 Terminologie der PKS

In letzter Zeit wurde verschiedentlich darauf hingewiesen, dass sich die Vielfalt bislang gefundener PKS nur schwer in ein wirklich konsistentes Klassifikationsschema einordnen lässt (Shen 2003; Müller 2004). Diese Schwierigkeiten beruhen hauptsächlich auf der Tendenz, strukturelle Gesichtspunkte mit funktionellen korrelieren zu wollen. Darüber hinaus treten Probleme auf, wenn versucht wird, diejenigen Organismengruppen, in denen eine bestimmte PKS-Variante scheinbar ausschließlich auftritt, in die Terminologie mit einzubeziehen. So ist der Begriff „pflanzliche PKS“, mit dem früher PKS-III belegt wurden, daraus entstanden, dass man annahm, sie kämen nur in Pflanzen vor, wären also ein originär pflanzliches Bio-synthese-System. Mit der Entdeckung von PKS-III in Bakterien und Pilzen (siehe Abschnitt 1.2.4) wurde diese Bezeichnung schließlich obsolet. Ebenso dürfen PKS-I nicht durchgehend mit modularen PKS gleichgesetzt werden, die nicht iterativ arbeiten, da neben diesen auch nicht modular aufgebaute PKS-I existieren, die tatsächlich iterative Funktionalität zeigen (siehe Abschnitte 1.2.2.1 und 1.2.2.2). Dass es innerhalb dieser Untergruppen wiederum zu Abweichungen im Funktionsprinzip kommt, wie es sich z. B. anhand der partiellen Iterativität in modular aufgebauten Synthasen (siehe Abschnitt 1.2.2.1.4) zeigen lässt, schränkt die grundsätzliche Klassifizierbarkeit nicht ein, insbesondere wenn man sich dabei auf die strukturelle Abgrenzung beschränkt, wie sie der Untergliederung in Typ I und Typ II zugrunde liegt, und diese durch die Sondergruppe eines Typs III ergänzt, der sich zwanglos von den anderen beiden unterscheiden lässt.

1.3 Nichtribosomale Peptidsynthetasen

In den 60er Jahren des 20. Jahrhunderts erkannte man, dass bestimmte kleine Peptid-Verbindungen nicht von Ribosomen synthetisiert werden. Eines der ersten Beispiele für solche nichtribosomalen Peptide war das als Ionophor wirkende Peptid-Antibiotikum Gramicidin S (Gevers, Kleinkauf und Lipmann 1968). In den folgenden Jahren wurden zum einen immer mehr Substanzen mit einem Peptid-Anteil identifiziert, deren Biosynthese nichtribosomal verläuft, zum anderen konnten die zugrunde liegenden Enzymsysteme gefunden und in ihrer Funktion aufgeklärt werden. Diese nichtribosomalen Peptidsynthetasen (NRPS) sind modular aufgebaut und zeigen ein Syntheseprinzip, das in vielerlei Hinsicht an dasjenige der modularen PKS-I erinnert (siehe Abschnitt 1.2.2.1). Sie kommen in Bakterien und Pilzen vor. Während in Pflanzen noch keinerlei NRPS-ähnliche Enzyme beschrieben worden sind, gibt es seit kurzem Hinweise darauf, dass sie auch in höheren Eukaryoten auftreten, wenngleich in einer im Vergleich zu Bakterien und Pilzen stark reduzierten Form (Kasahara und Kato 2003; Richardt *et al.* 2003). Charakteristische Merkmale von

nichtribosomal hergestellten Peptiden sind das Auftreten nichtproteinogener Aminosäuren, darunter auch solche mit D-Konfiguration am α -C-Atom, und die Ringbildung nach Abschluss des Aminosäure-Einbaus.

1.3.1 Aufbau von NRPS und allgemeines Synthese-Schema

NRPS sind insofern modular aufgebaut, als jeder Schritt der Biosynthese von einer gleichartig organisierten Einheit von enzymatischen Aktivitäten durchgeführt wird, die hintereinander geschaltet sind. Dabei sind diese Aktivitäten auf umschriebenen Domänen innerhalb der Polypeptid-Kette lokalisiert. Jedes Modul ist für den Einbau eines Aminosäure-Restes zuständig. Die Domänen-Typen lassen sich auf der Ebene der Aminosäure-Sequenz eindeutig anhand von charakteristischen, stark konservierten Sequenz-Motiven, den so genannten „Kern-Motiven“ (englisch: *core motifs*) erkennen. Als Beispiel dient hier der Multienzymkomplex, der das Peptid-Gerüst von Bacitracin A in *Bacillus licheniformis* synthetisiert (Konz *et al.* 1997). Das Synthese-Schema ist in **Abb. 7** dargestellt. Die etwa 550 Aminosäure-Reste umfassende Adenylierungs-Domäne (A-Domäne) aktiviert die neu einzubauende Aminosäure über eine Aminoacyl-Zwischenstufe unter Verbrauch von ATP und überträgt sie auf die Thiolierungs-Domäne (T-Domäne), auch PCP genannt (von englisch: *peptidyl carrier protein*), an die sie in Form eines Thioesters kovalent gebunden ist. T-Domänen sind typischerweise zwischen 80 und 100 Aminosäuren lang. Die Aktivierung der T-Domäne erfolgt wie im Fall der ACP-Domänen von PKS posttranslational durch Phosphopantetheinyl-Transferasen (Gocht und Marahiel 1994; Reuter *et al.* 1999; Mootz, Finking und Marahiel 2001). Die Kondensations-Domäne (C-Domäne, von englisch: *condensation domain*), die mit ca. 450 Aminosäureresten ähnlich groß wie die A-Domäne ist, knüpft eine neue Peptid-Bindung zwischen dem Peptidyl-Rest eines Moduls und dem Aminoacyl-Rest des nachfolgenden Moduls. Dieser Prozess wiederholt sich in Form eines Fließband-Ablaufs linear entlang der gesamten NRPS-Anordnung. Nach Ausbildung der letzten Bindung wird die Peptid-Kette durch eine Thioesterase-Domäne (TE-Domäne) vom Enzymkomplex abgespalten, wobei im Fall der Surfactin-Synthetase gleichzeitig eine Zyklisierung erfolgt.

Ähnlich wie bei den modularen PKS-I gilt für viele NRPS die Colinearitätsregel, wonach sich die Reihenfolge der Aminosäure-Reste im Peptid direkt aus der Abfolge der Module innerhalb des NRPS-Komplexes ergibt. In Bakterien sind die Module üblicherweise auf mehrere Proteine verteilt, die sich zu einem Multienzymkomplex zusammenlagern, wobei die Zahl der Module pro Protein variiert. Eines der größten bisher charakterisierten NRPS-Proteine ist die Syringomycin-Synthetase E aus *Pseudomonas syringae*, in der acht Module vereinigt sind und die bei einer Länge von 9377 Aminosäure-Resten eine molekulare Masse von etwa 1 MDa besitzt (Guenzi *et al.* 1998). In Pilzen sind alle zur Biosynthese einer Verbindung notwendigen Module häufig auf einem einzigen Protein integriert, wodurch ihre NRPS extreme Größen erreichen. Die Cyclosporin A-Synthetase aus *Tolypocladium niveum* besteht aus 11 Modulen und weist eine molekulare Masse von 1,6 MDa auf (Weber *et al.* 1994).

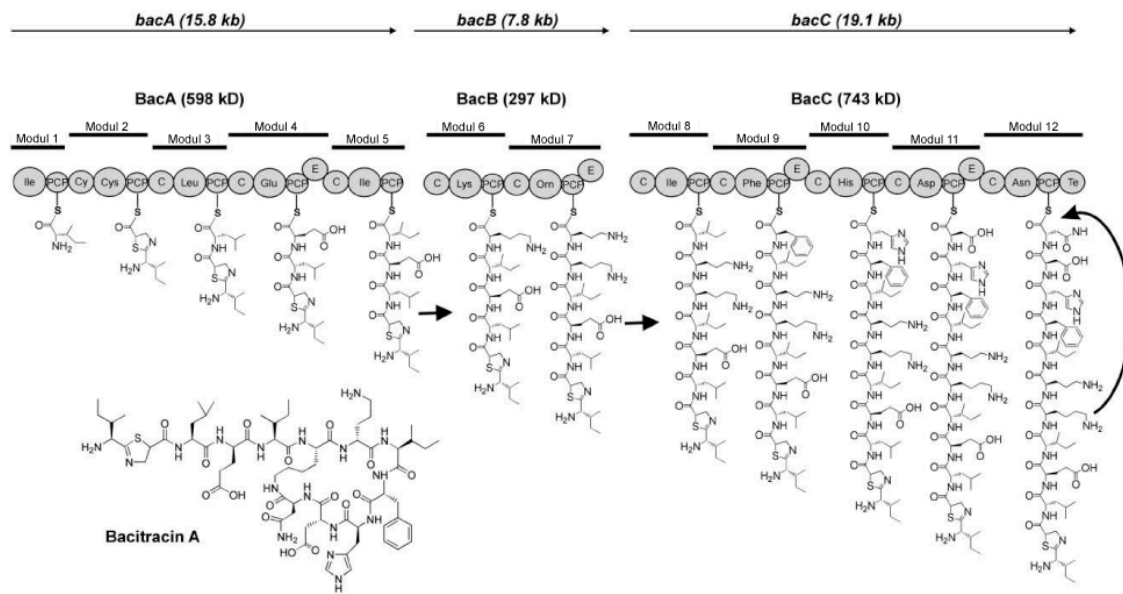


Abb. 7: Bacitracin-Biosynthese in *B. licheniformis*.

Die NRPS ist aus 12 Modulen aufgebaut, die hintereinander geschaltet jeweils einen Aminosäure-Baustein (angegeben durch die Dreibuchstaben-Abkürzung) in das Peptid einbauen. (Nach Schwarzer, Finking und Marahiel 2003, verändert.)

1.3.2 Enzymologie und Struktur der NRPS-Domänen

1.3.2.1 Adenylierungs-Domänen

A-Domänen aktivieren die einzubauenden Aminosäuren mit einem Zwei-Schritt-Mechanismus (Marahiel, Stachelhaus und Mootz 1997). Im ersten Schritt wird die Aminosäure an das Enzym gebunden und mit ATP unter Freisetzung von Pyrophosphat (PP_i) in das entsprechende Aminoacyl-Adenylat umgewandelt. Im zweiten Schritt wird dann durch einen nucleophilen Angriff der freien Thiolgruppe des Phosphopantetheinyl-Arms der T-Domäne dieser Oxoester in einen Thioester überführt, wobei AMP freigesetzt wird. Dieser Ablauf erinnert an die Beladung von tRNA durch Aminoacyl-tRNA-Synthetasen, wenngleich in diesem Fall die Aminosäure an eine Hydroxyl-Gruppe des 3'-Endes der tRNA und somit als Oxoester gebunden wird. Trotz dieser mechanistischen Übereinstimmungen zeigen A-Domänen weder in ihrer Sequenz noch strukturell besondere Ähnlichkeiten mit Aminoacyl-tRNA-Synthetasen (Arnez und Moras 1997). Sie gehören vielmehr zu einer Familie Adenylat-bildender Enzyme, die unter anderem die aus dem Primärstoffwechsel bekannten Acyl-CoA-Ligasen und die Luciferase enthält (Turgay, Krause und Marahiel 1992; Weber und Marahiel 2001). Aus der Kristallstruktur von PheA, dem Phenylalanin-aktivierenden ersten Modul der Gramicidin S-Synthetase, ergaben sich wichtige Aufschlüsse hinsichtlich der strukturellen Organisation (Conti *et al.* 1997). Danach wird die A-Domäne aus einer kleineren, etwa 100 Aminosäure-Reste umfassenden C-terminalen Subdomäne und einer größeren, ungefähr 400

Aminosäure-Reste langen N-terminalen Subdomäne aufgebaut. Das aktive Zentrum ist an der Verbindungsfläche zwischen diesen Subdomänen lokalisiert, wobei fast alle für die Substraterkennung wichtigen Aminosäure-Positionen Teil der C-terminalen Domäne sind. Aus diesen Strukturdaten in Verbindung mit den Aminosäure-Sequenzen charakterisierter A-Domänen konnten Regeln für die Vorhersage der Substratspezifität abgeleitet werden (siehe Abschnitt 1.3.3). Im Gegensatz zur ribosomalen Proteinsynthese, bei der die Auswahl der korrekten tRNA mit großer Genauigkeit erfolgt und durch einen Korrekturlese-Mechanismus ergänzt wird, findet man bei A-Domänen eine gewisse Flexibilität hinsichtlich der Substratauswahl (Stachelhaus, Mootz und Marahiel 1999), so dass dieselbe NRPS diverse Produkt-Strukturvarianten produzieren kann. Das cyclische Dekapeptid Tyrocidin besteht beispielsweise aus vier verschiedenen Verbindungen, die sich untereinander in zwei Aminosäure-Positionen unterscheiden (Ruttenberg und Mach 1966). Das immunosuppressiv wirkende Cyclosporin weist sogar ca. 30 Strukturvarianten auf (Weber *et al.* 1994). Neuere Arbeiten haben allerdings gezeigt, dass A-Domänen über eine intrinsische ATPase-Aktivität verfügen, die in Gegenwart nicht passender Aminosäuren ansteigt und im Fall von Domänen mit bekannter relaxierter Substratauswahl deutlich geringer ausfällt (Pavela-Vrancic, Dieckmann und von Döhren 2004).

1.3.2.2 Kondensations-Domänen und Cyclisierungs-Domänen

C-Domänen knüpfen die Peptidbindung zwischen dem Peptidyl-Donor des vorausgehenden Moduls und dem Aminoacyl-Akzeptor des eigenen Moduls und sorgen damit gleichzeitig für die Translokation der verlängerten Peptid-Kette innerhalb des NRPS-Multienzymkomplexes (Marahiel, Stachelhaus und Mootz 1997). Diese Funktionen konnten durch die Einführung von Deletionen nachgewiesen werden (Stachelhaus, Huser und Marahiel 1996). Da bislang keine Strukturinformationen zu diesem Domänentyp erhalten werden konnten, ist der katalytische Mechanismus der Bildung der Peptidbindung immer noch ungeklärt, ebenso fehlen Informationen darüber, wie sie mit den anderen Domänen desselben Moduls und der Nachbar-Module interagiert. Hingegen konnte biochemisch gezeigt werden, dass auch C-Domänen Selektivität gegenüber ihren Substraten zeigen, wobei die Selektivität hinsichtlich des neu einzubauenden Aminoacyl-Restes deutlich höher ausgeprägt ist als hinsichtlich des von dem stromaufwärts gelegenen Modul bereit gestellten Peptidyl-Restes (Belshaw, Walsh und Stachelhaus 1999). Die Unterscheidungsfähigkeit der C-Domänen scheint im Wesentlichen in der Größe der Seitenkette und den stereochemischen Eigenschaften des Aminoacyl-Restes begründet zu sein (Linne und Marahiel 2000).

Es existiert eine Variante von C-Domänen, die neben der Bildung der Peptidbindung auch die Heterozyklisierung des neuen Aminoacyl-Restes zu katalysieren vermag (Sieber und Marahiel 2005). Dabei kommt es im Fall von Serin und Threonin zur Ausbildung 5-gliedriger Oxazolin-Ringe, während im Fall von Cystein 5-gliedrige Thiazolin-Ringe entstehen. Durch sich anschließende Reaktionen können diese Heterozyklen weiter modifiziert werden. So führt bei der Biosynthese von Pyochelin in *Pseudomonas aeruginosa* eine in *trans* arbeitende NADPH-abhängige Reduktase zu einer Thiazolidin-Struktur (Reimann *et al.* 2001). Im Verlauf der Biosynthese von Epothilon wird hingegen das Thiazolin durch eine Oxidase-Domäne (Ox-Domäne), die Bestandteil der A-Domäne der NRPS-Komponente EpoB ist, mittels des Transfers von zwei Elektronen auf die Stufe eines Thiazol-Rings oxidiert (Schneider, Shen und Walsh 2003).

1.3.3 Selektion der Aminosäuren durch die A-Domänen

Auf Basis der Kristallstruktur der Phenylalanin-aktivierenden A-Domäne der NRPS PheA, die zusammen mit L-Phenylalanin und AMP kristallisiert worden war, konnte die Bindungstasche für das Substrat ermittelt werden (Conti *et al.* 1997). In dieser Bindungstasche stehen 10 Aminosäure-Reste in direktem Kontakt mit Phenylalanin. Mit Hilfe eines Aminosäure-Alignments des relevanten Teils von PheA mit etwa 150 anderen, hinsichtlich ihrer Substratspezifität charakterisierten A-Domänen wurden in zwei unabhängigen Untersuchungen so genannte Signatur-Sequenzen ermittelt (Stachelhaus, Mootz und Marahiel 1999; Challis, Ravel und Townsend 2000). Diese Signaturen geben einen empirischen Zusammenhang zwischen den Aminosäure-Resten in diesen Positionen der Substrat-Bindungstasche und der aktivierten Aminosäure an, entsprechen also gewissermaßen einem Spezifitäts-Code der A-Domänen. Mit diesem Code sollte es theoretisch möglich sein, die Spezifität einer A-Domäne allein aufgrund ihrer Aminosäure-Sequenz zu ermitteln. Darüber hinaus eröffnet diese Zuordnung das Potenzial, die Spezifität von A-Domänen mittels positionsspezifisch eingeführter Mutationen gezielt zu verändern. Zwei der 10 Positionen erweisen sich bei näherer Betrachtung als wenig aussagefähig, da sie praktisch vollständig konserviert sind (Stachelhaus, Mootz und Marahiel 1999; Challis, Ravel und Townsend 2000). In der einen Position findet sich fast immer ein Aspartat-Rest, der mit der Amino-Gruppe der Substrat-Aminosäure wechselwirkt, die andere Position wird von einem Lysin-Rest besetzt, dessen Seitenkette in Wechselwirkung mit der Carbonssäure-Gruppe des Substrats steht (Conti *et al.* 1997). Daher verbleiben noch acht Positionen zur Spezifitäts-Vorhersage. Im Ergebnis konnte festgestellt werden, dass der auf die Bindungstasche bezogene Spezifitäts-Code für bestimmte Typen von Aminosäuren eindeutige Ergebnisse liefert, bei anderen Typen hingegen große Unsicherheiten verbleiben. So zeigen insbesondere Aminosäuren mit polaren Seitenketten und Cystein eine klare Korrelation, während die Situation bei unpolaren Aminosäuren weniger eindeutig ist (Challis, Ravel und Townsend 2000).

Ein Grundproblem dieser Methode zur Ermittlung eines Spezifitäts-Codes besteht darin, dass die Bindungstasche von PheA, die als Substrat Phenylalanin mit seiner langen Seitenkette aufnimmt, ein schlechtes Modell für Aminosäuren mit vergleichsweise kurzen Seitenketten ist (Lautru und Challis 2004). Eine Modellierungs-Studie legte z. B. den Schluss nahe, die Erkennung von L-Threonin wäre möglicherweise durch nur zwei Kontakte in der Bindungstasche bestimmt (Ackerley, Caradoc-Davies und Lamont 2003). Weitere Probleme sind in einem kürzlich erschienen Übersichtsartikel zusammengefasst (Lautru und Challis 2004).

1.3.4 Modifizierende Enzymaktivitäten in NRPS

Neben den Domänen-Typen, die in jedem NRPS-Modul vorkommen, gibt es noch akzessorische Domänen, die eine Modifizierung der gebundenen Aminoacyl-Reste bewirken, bevor diese in das Peptid eingebaut werden. Dabei handelt es sich zum einen um Epimerisierungs-Domänen (E-Domänen), die eine Konfigurationsumkehr an der asymmetrischen α -Position des Aminoacyl-Restes katalysieren, zum anderen um Methylierungs-Domänen (M-Domänen), die Methyl-Gruppen in das Peptid einführen. E-Domänen sind etwa 450 Aminosäure-Reste lang und scheinen neben der Bereitstellung einer Epimerase-

Aktivität in manchen NRPS auch eine wichtige Rolle bei der Erkennung der korrekten Protein-Protein-Wechselwirkungen innerhalb des Biosynthese-Komplexes zu spielen (Linne *et al.* 2003). M-Domänen übertragen Methyl-Gruppen von S-Adenosylmethionin (SAM) auf bestimmte Positionen der Peptidkette. Dieser Domänen-Typ umfasst etwa 420 Aminosäure-Reste und ist in die A-Domäne desselben Moduls integriert (Sieber und Marahiel 2005). Am häufigsten findet man N-methylierte Peptidbindungen, daneben aber auch Methylierungen in anderen Positionen, so z. B. in Yersiniabactin an einem C-Atom des Thiazolin-Rings (Miller, Walsh und Luo 2001).

1.4 Gemischte Biosynthese mit PKS- und NRPS-Modulen

Neben modularen Biosynthese-Systemen, die ausschließlich aus PKS- oder NRPS-Modulen bestehen, existieren auch gemischte Varianten, in denen beide Modultypen vorkommen. Im Hinblick auf die Anordnung der kooperierenden Module muss zwischen PKS/NRPS-Konfigurationen und NRPS/PKS-Konfigurationen unterschieden werden. Bei den ersteren wird ein NRPS-Modul von einem vorgelagerten PKS-Modul bedient, bei den letzteren liegen die Verhältnisse umgekehrt (Shen *et al.* 2001). Solche hybriden Systeme sind beispielsweise die Rapamycin-Synthase, die aus 12 PKS-Modulen und einem NRPS-Modul besteht (Aparicio *et al.* 1996), und die Yersiniabactin-Synthetase, in der drei NRPS-Module und ein PKS-Modul zusammen arbeiten (Gehring *et al.* 1998). In diesen Fällen liegen die Module der beiden unterschiedlichen Enzym-Familien auf getrennten Polypeptid-Ketten. Daneben sind auch Biosynthese-Komplexe beschrieben worden, in denen NRPS- und PKS-Module Bestandteile derselben Polypeptid-Kette sind. Zu dieser Klasse gehört das erste Modul in der Synthese des Antibiotikums TA in *Myxococcus xanthus* (Paitan *et al.* 1999).

1.5 Biotechnologisches Potenzial von modularen PKS-I und NRPS

Wie schon bald nach der ersten Charakterisierung modularer PKS-I erkannt wurde, birgt der Biosynthese-Mechanismus modular aufgebauter Enzymsysteme die Möglichkeit zur gezielten Herstellung neuer Verbindungen (Hutchinson und Fujii 1995). Dahinter steht die Idee, dass sich durch genetische Manipulation der codierenden DNA-Sequenzen unter Ausnutzung des modularen Prinzips neue Modul-Kombinationen generieren lassen sollten, die dementsprechend gegenüber den ursprünglichen Metaboliten zu veränderten Strukturen führen. Da bei dieser Technik natürlich vorkommende Biosynthese-Enzyme zum Einsatz kommen, deren Komponenten gentechnisch zu neuen Kombinationen zusammengestellt werden, beschreibt man diese experimentellen Ansätze zusammenfassend mit dem Begriff „kombinatorische Biosynthesen“. Insbesondere auf der Basis charakterisierter modularer PKS-I aus Streptomyceten wurden in den vergangenen 10 Jahren zahlreiche Versuche unternommen, derartige Techniken zu entwickeln, die in neueren Übersichtsartikeln in allen Aspekten beschrieben werden (Staunton und Weissman 2001; Reeves 2003). Es wurden verschiedene Arten von Modifikationen

eingesetzt, wobei zwei Herangehensweisen zu unterscheiden sind. Einerseits wird versucht, einzelne Domänen durch andere zu ersetzen, wodurch im Fall der AT-Domänen eine Änderung der Substratspezifität und damit eine Veränderung der Alkyl-Ketten des Polyketid-Grundgerüsts erreicht werden kann, oder Domänen werden inaktiviert oder vollständig aus einem Modul entfernt, so dass sich die Reduktionsstufe des Polyketids ändert. In diesen Fällen werden also Veränderungen in den Modulen herbeigeführt. Zum anderen lässt sich die Strategie verfolgen, die im natürlichen System vorhandenen Module selbst unverändert zu lassen und stattdessen die Anordnung der Module in der Biosynthese-Kette zu modifizieren, was gemäß der Colinearitätsregel (siehe Abschnitt 1.2.2.1.1) ebenfalls zu neuen Strukturen führt. Obwohl die Möglichkeit zu kombinatorischen Biosynthesen mit beiden Herangehensweisen experimentell erfolgreich demonstriert werden konnte, zeigte sich eine Reihe von praktischen Schwierigkeiten. Insbesondere wurde immer wieder beobachtet, dass die gentechnisch veränderten Synthasen weitaus weniger effizient Produkte herstellen als die Wildtypen (Staunton und Wilkinson 2001). Weiterhin erweisen sich die natürlichen Produzenten-Stämme oft als wenig geeignet für genetische Manipulationen, weshalb es im Allgemeinen sehr zeitaufwendig ist, eine größere Zahl an Neukombinationen herzustellen und zu testen (Pfeifer und Khosla 2001).

Gentechnische Ansätze zur Veränderung von NRPS ähneln prinzipiell denen bei modularen PKS-I (Staunton und Wilkinson 2001). So konnten neue Surfactin-Varianten mit anderen Aminosäure-Resten durch den Austausch von Einheiten aus A- und T-Domänen erzeugt werden, allerdings wurde in allen Fällen eine sehr geringe Produktion von höchstens 0,5 % der Menge des normalen Surfactins im Wildtyp gefunden (Stachelhaus, Schneider und Marahiel 1995). Im Fall von NRPS besteht eine zusätzliche Strategie darin, keine größeren Veränderungen vorzunehmen, um die Interaktionen innerhalb des Enzymkomplexes nicht zu stören, sondern durch einige wenige Punktmutationen die Substratspezifität von A-Domänen zu ändern und sich dabei an den Kenntnissen über den Spezifitäts-Code (siehe Abschnitt 1.3.3) zu orientieren. Auf diese Weise konnten neue Peptide hergestellt werden, mit zufrieden stellender Effizienz jedoch nur, wenn die Substrat-Änderung chemisch ähnliche Aminosäuren betraf wie Glutamin gegen Glutamat oder Asparagin gegen Aspartat (Eppelmann, Stachelhaus und Marahiel 2002). Trotz gewisser Fortschritte, ist man von echter Biokombinatorik, die viele neue Verbindungen in relevanter Menge synthetisieren können sollte, noch weit entfernt (Sieber und Marahiel 2005).

1.6 Cyanobakterielle Sekundärmetabolite des PKS- und NRPS-Typs

1.6.1 Übersicht

Die in den vorher gehenden Abschnitten beschriebenen PKS- und NRPS-Biosynthese-Systeme kommen auch in der Gruppe der Cyanobakterien vor. Cyanobakterien sind photolithotrophe Prokaryoten, die mittels Photosynthese CO₂ assimilieren und dabei Wasser als Elektronendonator benutzen (Gottschalk 1986). Sie sind die einzige Bakteriengruppe, die zur oxygenen Photosynthese befähigt ist. Sie kommen in allen Lebensräumen sowohl aquatisch als auch terrestrisch vor und besiedeln auch ökologische Nischen wie heiße Quellen und extrem trockene Standorte (Whitton und Potts 2000). Aufgrund ihrer Photosynthesefähigkeit gehören sie zu den wichtigen Primärproduzenten (Liu, Campbell und Nolla 1997). Außerdem sind manche Arten zur Fixierung molekularen Stickstoffs befähigt, wofür sie spezielle Zellformen, die so

genannten Heterocysten, ausbilden. Diese diazotrophen Formen spielen eine wichtige Rolle im globalen N₂-Haushalt (Montoya *et al.* 2004). Cyanobakterien sind eine sehr alte Gruppe innerhalb der Prokaryoten. In 2,7 Milliarden Jahren altem Schieferstein konnten die für Cyanobakterien charakteristischen 2-Methylhopanoide gefunden werden, die als „molekulare Fossilien“ aufgefasst und als Beleg für das Alter dieser Gruppe angeführt wurden (Brocks *et al.* 1999), wenngleich diese Schlussfolgerungen nicht ausreichend abgesichert zu sein scheinen (Knoll 2003). Außer Frage steht jedoch, dass es sich um eine evolutionär sehr alte Bakteriengruppe handelt.

Aufgrund morphologischer Kriterien werden Cyanobakterien traditionell in fünf Untergruppen eingeteilt (Rippka *et al.* 1979): Gruppe I umfasst solche einzellige Formen, die sich ausschließlich durch binäre Spaltung oder Knospung vermehren. In Gruppe II sind Cyanobakterien zusammengefasst, die einzellig sind und sich auch durch multiple Spaltung vermehren können. Gruppe III beschreibt mehrzellige, fädige Cyanobakterien, die keine Zelldifferenzierungen zeigen, also auch keine Heterocysten ausbilden können. Die Vertreter der Gruppe IV treten ebenfalls in mehrzelligen Fäden auf und sind im Unterschied zur Gruppe III zur Bildung von Heterocysten befähigt. Die Angehörigen der Gruppe V schließlich unterscheiden sich von denen der Gruppe IV durch Zellteilungen in mehr als einer Ebene.

Aus Cyanobakterien wurde bereits eine Reihe von Sekundärmetaboliten isoliert, die zu den Gruppen der Polyketide und nichtribosomalen Peptide gehören. In den letzten zwanzig Jahren wurden allein mehr als 600 Peptide und Metabolite mit Peptid-Anteil beschrieben (Welker und von Döhren 2006). Mit molekular-genetischen Ansätzen konnte gezeigt werden, dass zahlreiche Cyanobakterien-Spezies für NRPS codierende Gene besitzen (Christiansen *et al.* 2001). Viele cyanobakterielle Sekundärmetabolite zeigen Bioaktivitäten, die sie aus medizinisch-pharmazeutischer Sicht interessant erscheinen lassen oder ökologisch bedeutsam sind (Dittmann, Neilan und Börner 2001). Beispielsweise wirkt das aus *Microcystis aeruginosa* stammende Cyanopeptolin als Trypsin-Inhibitor (Jakobi *et al.* 1995). Die tropische Spezies *Lyngbya majuscula* produziert das Mitose-Gift Curacin A (Blokhin *et al.* 1995), dessen Biosynthese-Cluster aus PKS- und NRPS-Komponenten besteht (Chang *et al.* 2004). Stämme derselben Spezies synthetisieren Barbamid, eine chlorierte Verbindung, die toxisch auf Mollusken wirkt und ebenfalls durch ein gemischte PKS/NRPS-Biosynthese gebildet wird (Orjala und Gerwick 1996; Chang *et al.* 2002). Die Microcystine aus *Microcystis aeruginosa* waren das erste Beispiel einer Biosynthese unter Beteiligung von NRPS in Cyanobakterien (Moore *et al.* 1991; Dittmann *et al.* 1997). Microcystine bilden eine Gruppe von mehr als 60 strukturell verwandten zyklischen Heptapeptiden. Sie sind Leber-toxisch, da sie spezifisch die eukaryotischen Protein-Phosphatasen 1 und 2A inhibieren, und wegen ihres weltweiten Vorkommens von ökologischer Bedeutung (Dittmann und Wiegand 2006).

Mit der steigenden Zahl an veröffentlichten bakteriellen Genom-Sequenzen ergeben sich neue Möglichkeiten der Untersuchung von Sekundärmetaboliten. Dies gilt gerade auch für Bakterien mit großen Genomen, die über vergleichsweise komplexe Eigenschaften verfügen. In diese Gruppe fällt die Cyanobakterien-Spezies *Nostoc punctiforme*.

1.6.2 Das Cyanobakterium *Nostoc punctiforme*

Das Cyanobakterium *N. punctiforme* gehört nach der im voran gegangenen Abschnitt angeführten Klassifikation der Gruppe IV an, kommt also in mehrzelligen Fäden vor und kann zum Zweck der N₂-Fixierung Heterocysten ausbilden. Das Genom von *N. punctiforme* ist mit 8,9 Mbp im Vergleich zu anderen Cyanobakterien sehr groß und das derzeit größte vollständig sequenzierte Cyanobakterien-Genom (Meeks *et al.* 2001). Insgesamt finden sich bei diesem Cyanobakterium vier unterschiedliche Zelltypen, womit es innerhalb des Reiches der Prokaryoten zu den Organismen mit der am höchsten entwickelten Differenzierungsfähigkeit gehört (Meeks *et al.* 2002). Außerdem ist *N. punctiforme* zur Ausbildung von symbiotischen Gemeinschaften mit vielen Pflanzen und Pilzen befähigt (Meeks und Elhai 2002).

Neben den vegetativen Zellen, die den unspezialisierten Grundzelltyp darstellen, treten Heterocysten, Akineten und Hormogonien auf. Die Umwandlung von vegetativen Zellen in Heterocysten wird durch einen Mangel an verwertbaren Stickstoff-Verbindungen ausgelöst (Adams 2000). An diesem Prozess, bestehend aus der Signalkette, welche die Differenzierung einleitet, dem strukturellen Umbau der Zelle und der Bereitstellung der notwendigen speziellen Enzymaktivitäten, insbesondere der Nitrogenase, sind zahlreiche Gene beteiligt. Neuere Schätzungen, die aus Arbeiten an *Anabaena* stammen, gehen davon aus, dass ihre Zahl bei mindestens 140 liegt (Wolk 2000). Da das Enzym Nitrogenase in Anwesenheit von Sauerstoff seine Aktivität verliert, sind die Differenzierungsvorgänge darauf ausgerichtet, den Sauerstoffgehalt im Cytoplasma möglichst stark zu reduzieren. Dies wird erreicht durch die Repression der Synthese von Enzymen, die an der Photosynthese und CO₂-Assimilierung beteiligt sind, wie die Phycobiliproteine, der O₂-bildende Komplex des Photosystems II und die Ribulosebisphosphat-Carboxylase, und durch den Aufbau einer extrem dicken Zellwand aus einer zweischichtigen Polysaccharid-Schicht und einer Glycolipid-Ummantelung, womit die Durchlässigkeit für Sauerstoff, Ionen und andere hydrophile Verbindungen stark eingeschränkt wird (Wolk, Ernst und Elhai 1994).

Akineten zeichnen sich gegenüber vegetativen Zellen durch eine deutlich größere Resistenz gegen Kälte und Trockenheit aus (Meeks *et al.* 2002). Es handelt sich jedoch nicht um Sporen-ähnliche Spezialisierungen, wie sie bei Gram-positiven Bakterien auftreten, sondern sie zeigen deutliche Stoffwechselaktivität (Thiel und Wolk 1983) und ähnliche Mengen an DNA, RNA und Proteinen wie vegetative Zellen (Sutherland, Herdman und Stewart 1979). Das primäre Signal zur Auslösung der Akineten-Entwicklung scheint eine allgemeine Energie-Limitierung zu sein (Meeks *et al.* 2002).

Hormogonien sind relativ kurze, zu einer Gleitbewegung befähigte Filamente, die zum einen eine Verbeitung des Organismus auf kurze Distanzen ermöglichen, zum anderen grundsätzlich die infektiösen Einheiten bei der Ausbildung symbiontischer Beziehungen darstellen (Meeks 1998). Vermutlich wird die Ausbildung von Hormogonien durch eine ganze Reihe von Signalen aus der Umwelt beeinflusst. Die genauen Mechanismen sind noch nicht bekannt, verschiedene Experimente haben jedoch unter anderem den Einfluss der Nährstoff- und Salzkonzentration (Herdman und Rippka 1988) und der Lichtverhältnisse (Damerval *et al.* 1991) aufgedeckt.

Aus den verschiedenen *Nostoc*-Spezies konnten bereits viele Sekundärmetabolite isoliert werden, darunter Aminosäure-Derivate, Terpenoide und spezielle Polysaccharide (Dembitsky und Rezanka 2005). Neben den Cryptophycinen (Golakoti *et al.* 1995; Magarvey *et al.* 2006) wurden auch andere Peptide

gefunden. Dazu gehören die Nostopeptolide A₁, A₂ und A₃ (Golakoti *et al.* 2000) und die Nostocyclopeptide A₁ und A₂ (Golakoti *et al.* 2001). Stämme der Gattung *Nostoc* gelten daher allgemein als wichtige Quelle von Sekundäretaboliten (Biondi *et al.* 2004). In Extrakten von *Nostoc punctiforme* konnten bislang die Nostopeptolide A₁ und A₃ identifiziert werden (Hunsucker *et al.* 2004). Mit der veröffentlichten Genom-Sequenz dieses Cyanobakteriums wird ein direktes Auffinden von PKS- und NRPS-Biosynthese-Clustern möglich, was neue Wege für die Identifizierung und Untersuchung der von diesen Clustern produzierten Sekundärmetabolite eröffnet.

1.7 Zielstellung

Sekundärmetabolite und die sie produzierenden Enzyme stehen in enger Wechselbeziehung zum Primärstoffwechsel. Dies gilt auch für die modularen Multienzymkomplexe des bakteriellen Sekundärstoffwechsels. Vereinzelt wurden zwar Arbeiten vorgelegt, die sich punktuell mit evolutionären Gesichtspunkten dieses Teils des Stoffwechsels befassen, es fehlen jedoch insbesondere im Hinblick auf PKS und NRPS umfassende Studien, welche die evolutionären Zusammenhänge darlegen. Die vorliegende Arbeit sollte diese Lücke für den Bereich der Bakterien schließen, wobei der Schwerpunkt auf die modularen PKS des Typs I gelegt wurde. Folgende Fragestellungen standen dabei im Vordergrund:

- In welcher evolutionären Beziehung stehen bakterielle modulare PKS-I im Gesamtzusammenhang von FAS und PKS aus allen Organismengruppen?
- Welche Verteilung zeigen diese Systeme in den verschiedenen Bakteriengruppen und welche allgemeinen Rückschlüsse lassen sich daraus ziehen?
- Lassen sich bestimmte Muster und Mechanismen in der Evolution modularer PKS-I erkennen?

Diese Fragen sollten mit bioinformatisch-phylogenetischen Methoden untersucht werden. Ein weiteres Ziel dieser Arbeit war es, die Biosynthese-Cluster dieses Typs in dem Cyanobakterium *Nostoc punctiforme* auf der Grundlage der bekannten Genom-Sequenz zu identifizieren und in Teilen experimentell zu charakterisieren. Schließlich sollten die gewonnen Erkenntnisse in den Gesamtkomplex der Evolution metabolischer Diversität eingeordnet werden.

2 Material und Methoden

2.1 Material

2.1.1 Bakterienstämme

2.1.1.1 *Nostoc punctiforme*

Für die Untersuchungen an dem Cyanobakterium *Nostoc punctiforme* wurde der Stamm *N. punctiforme* ATCC 29133 der „American Tissue and Culture Collection“ verwendet. Die Zellen wurden von Prof. J. C. Meeks (University of California, Davis) zur Verfügung gestellt. Dieser Stamm entspricht dem in der „Pasteur Culture Collection“ hinterlegten Stamm *N. punctiforme* PCC 73102.

2.1.1.2 *Escherichia coli*

Für Klonierungen wurde der Stamm *E. coli* XL-1 blue (Stratagene, Amsterdam, Niederlande) eingesetzt, für heterologe Proteinexpressionen der Stamm *E. coli* Top 10 (Invitrogen, Groningen, Niederlande).

Für das *Triparental mating*-Protokoll zur Mutagenisierung von *N. punctiforme* wurde der *E. coli*-Stamm JCM113 verwendet (zur Verfügung gestellt von Prof. J. C. Meeks, University of California, Davis). Dieser Stamm basiert auf *E. coli* HB 101 und enthält das konjugative Plasmid pRK2013 (Figurski und Helinski 1979), das alle Komponenten des Konjugationsapparates codiert, und das Helfer-Plasmid pRL528 (Elhai und Wolk 1988). Letzteres stellt das Gen *mob* für ein *nicking*-Enzym zum Schneiden der *bom*-Stelle bereit und codiert außerdem für Methylasen, welche die in Cyanobakterien häufig vorkommenden Aval- und Avall-Schnittstellen methylieren.

2.1.2 Chemikalien

3-Morpholino-Propansulfonsäure (MOPS)	C. Roth, Karlsruhe
Agarose DNA	Biozym Diagnostik, Hameln
Acetonitril „HPLC Gradient Grade“	C. Roth, Karlsruhe
Aminosäuren	CalBiochem/Merck, Darmstadt
Ampicillin	Roche Diagnostics, Mannheim
L-Arabinose	SERVA Feinbiochemika, Heidelberg
Bacto-Agar	Difco, Detroit, USA
Bacto-Trypton	Difco, Detroit, USA
Bromphenolblau	SERVA Feinbiochemika, Heidelberg
Chloramphenicol	Roche Diagnostics, Mannheim
Chloroform	C. Roth, Karlsruhe

Hefeextrakt	Difco, Detroit, USA
Kanamycin	Roche Diagnostics, Mannheim
Natrium-Dodecylsulfat (SDS)	SERVA Feinbiochemika, Heidelberg
Szintillationszählflüssigkeit Rotiscint Eco Plus	C. Roth, Karlsruhe
Tetranatrium-Pyrophosphat	Sigma-Aldrich, München

Alle weiteren Standard-Chemikalien und Lösungsmittel wurden von den Firmen C. Roth, Karlsruhe, Sigma-Aldrich, München oder SERVA Feinbiochemika, Heidelberg in den üblichen Reinheitsgraden bezogen.

2.1.3 Radiochemikalien

[γ - ³² P]-ATP, spezifische Aktivität 110 Tbq/mmol	Amersham Pharmacia Biotech, Freiburg
Tetranatrium-[³² -P]-Pyrophosphat	Perkin Elmer, Waltham, USA

2.1.4 Enzyme

CIAP (<i>calf intestine alkaline phosphatase</i>)	MBI Fermentas, St. Leon-Roth
Klenow-Fragment	MBI Fermentas, St. Leon-Roth
Lysozym	GERMED, Berlin
Pfu-Polymerase	Promega, Mannheim
Proteinase K	MBI Fermentas, St. Leon-Roth
Restriktionsendonucleasen	MBI Fermentas, St. Leon-Roth New England Biolabs, Schwalbach
RNAse A/T1-Mix	MBI Fermentas, St. Leon-Roth
T4-DNA-Ligase	MBI Fermentas, St. Leon-Roth
Tag-Polymerase	Qiagen, Hilden

2.1.5 Nucleinsäuren

Lambda-DNA (zur Marker-Herstellung)	MBI Fermentas, St. Leon-Roth
Oligonucleotide	Eurogentec, Köln Sigma-Genosys, Steinheim Metabion, Martinsried

Die für Klonierungen und Expressionen verwendeten Plasmide sind **Tab. 2** aufgelistet, die bei PCR-Ansätzen benutzten Primer in **Tab. 3**.

Tab. 2: Liste der verwendeten Plasmide

Plasmid	Antibiotika-Resistenzen gegen	Bezugsquelle oder Referenz
pGEM-T	Ampicillin	Promega, Mannheim
pBAD-TOPO	Ampicillin	Invitrogen, Karlsruhe
pRL271	Chloramphenicol, Erythromycin	(Black, Cai und Wolk 1993)
pUC4K	Ampicillin, Kanamycin	Amersham Pharmacia, Freiburg

Tab. 3: Liste der eingesetzten PCR-Primer

Name des Primers	Sequenz	Verwendung
NpF3173Exp-FWD	5'-GGAGGATGGCGAACTATGA	Expression von NpF3173
NpF3173Exp-REV	5'-TTTCCAGTTGTTCATAACTGCC	Expression von NpF3173
636/8Exp-FWD	5'-TTGAGTTTGCGGAAGATTTCT	Expression von NpF3179
636/8Exp-REV	5'-TCTGGCTAATCCATTCAGCAA	Expression von NpF3179
NpF3173KO-FWD	5'-GCATGCATGAGCGCGAGTCAACTTCT	Herstellung Knockout-Mutante
NpF3173KO-REV	5'-GAGCTCGAGACGCGGTGTGTAGGATT	Herstellung Knockout-Mutante
NpF3173KOcheck1	5'-CAAAACACATCAAGGTCAGAG	Überprüfung Knockout-Mutante
NpF3173KOcheck2	5'-TCTCTTTTTAGGCATCCCACTT	Überprüfung Knockout-Mutante
Reg-Bind-FWD	5'-GCAAAGTTAAGAATTAGCCAAGG	DNA mit Bindungsstelle
Reg-Bind-REV	5'-GCAAGCTTTTAATAATGCATACGA	DNA mit Bindungsstelle
NonReg-Bind-FWD	5'-TGCCATATCTTCGATAGTGGTG	DNA ohne Bindungsstelle
NonReg-Bind-REV	5'-TCATCCAGCTGGTGGAATG	DNA ohne Bindungsstelle
KM1800-FWD	5'-GCTAACTCACATTAATTGCGTTGC	Überprüfung Knockout-Mutante
KM375-REV	5'-CGT TTT ACA ACG TCG TGA CTG G	Überprüfung Knockout-Mutante

2.1.6 Kits

Qiagen „Gel Extraction Kit“	Qiagen, Hilden
Qiagen „DNA Polymerase PCR Kit“	Qiagen, Hilden
„Qiaquick PCR Purification Kit“	Qiagen, Hilden
Nickel-NTA-Superflow	Qiagen, Hilden
Bio-Rad Protein Assay	Biorad, München

2.1.7 Filter und Membrane

3MM-Filterpapier	Whatman, Maidstone, UK
HATF-Nitrocellulose-Membrane	Millipore, Bedford, USA
PVDF-Blotting-Membrane Hybond C	Amersham Pharmacia, Freiburg

2.1.8 Sequenzdaten

Alle für die phylogenetischen Analysen herangezogenen Sequenzdaten wurden aus öffentlich zugänglichen Datenbanken gewonnen. Die Datensuche erfolgte über den Server des *National Center for Biotechnology Information* (NCBI) mittels der dort zu Verfügung gestellten Blast-Suchfunktionen. Die Nucleotid- und Aminosäure-Sequenzen der PKS-Cluster von *Streptomyces avermitilis* stammen von der Website des *S. avermitilis*-Genomprojekts (<http://avermitilis.ls.kitasato-u.ac.jp>). Die DNA-Sequenzen der kleinen ribosomalen Untereinheit (16S-DNA) verschiedener Bakterien-Spezies stammen aus der *European ribosomal RNA database* (<http://www.psb.ugent.be/rRNA/ssu>).

2.2 Methoden

2.2.1 Mikrobiologische Methoden

2.2.1.1 Kultivierungsbedingungen für Bakterien

Zellen des Bakteriums *E. coli* wurden im Allgemeinen nach den üblichen Standardmethoden kultiviert (Sambrook und Russell 2001). Die Anzucht erfolgte in Luria-Bertani-Medium (LB-Medium), bestehend aus einer wässrigen Lösung von 1 % (w/v) Trypton, 0,5 % (w/v) Hefextrakt und 1 % (w/v) NaCl, die mittels Natronlauge auf einen pH-Wert von ca. 7,5 eingestellt war, entweder als Flüssigkultur auf einem Schüttler (220 rpm) bei 37° C oder unter Zusatz von 1 % (w/v) Bacto-Agar auf Kulturplatten ebenfalls bei 37° C. In Abhängigkeit der in die Zellen eingebrachten Plasmide wurde das LB-Medium zu Selektionszwecken mit Antibiotika der folgenden Endkonzentrationen versetzt: Amicillin (50 bis 100 µg/ml), Chloramphenicol (25 µg/ml) und Kanamycin (30 bis 40 µg/ml). In Kombination wurden Chloramphenicol und Kanamycin in den Endkonzentrationen 25 µg/ml bzw. 30 µg/ml zugesetzt. Im Fall der heterologen Expression von Proteinen wurden die Wachstumsbedingungen zur Optimierung der Expression angepasst. Die Zellen wurden zunächst bei 37° C bis zur frühen logarithmischen Phase angezogen, entsprechend einer optischen Dichte von etwa OD₆₀₀=0,6 bis 0,8. Nach der Induktion der Expression wurden die Zellen bei 25° C für vier Stunden weiter inkubiert.

Das Cyanobakterium *N. punctiforme* wurde standardmäßig in Allen&Arnon-Medium (AA-Medium) angezogen (Castenholz 1988), das in Flüssigkulturen als vierfache Verdünnung (AA/4) bereitgestellt wurde (Cohen *et al.* 1994). Für die Anzucht auf Kulturplatten wurde das AA-Medium mit Agar verfestigt. Zur Supplementierung mit gebundenem Stickstoff wurde das Medium mit NH₄Cl (Endkonzentration: 2,5 mM) versetzt.

2.2.1.2 Herstellung chemisch kompetenter *E. coli*-Zellen

200 ml LB-Medium, supplementiert mit 20 mM Glucose und 20 mM MgSO₄ und mit dem entsprechenden Antibiotikum versetzt, wurden mit einer Übernachtskultur im Verhältnis 1:100 inokuliert und bei 37° C bis zu einer Zelldichte von etwa 5×10^7 Zellen/ml angezogen. Die Zellen wurden durch 10-minütiges Zentrifugieren bei 4° C und 7.000 g geerntet und das Pellet in 100 ml eisgekühlter 50 mM CaCl₂-Lösung mittels vorsichtigem Pipettieren aufgenommen. Die Zellsuspension wurde für 60 min auf Eis inkubiert und dann erneut bei 4° C und 7.000 g zentrifugiert. Das Zellpellet wurde in 15 ml eisgekühlten FB-Puffer aufgenommen, in Aliquots aufgeteilt, mit flüssigem Stickstoff schockgefroren und bis zur Verwendung bei -70° C aufbewahrt.

2.2.1.3 Triparental mating zur Herstellung von *N. punctiforme*-Mutanten

Das angewandte *Triparental mating*-Verfahren („Drei-Eltern-Kreuzung“) basierte auf dem Mitte der 80er Jahre erstmals bei Cyanobakterien eingesetzten Originalverfahren (Wolk *et al.* 1984), wurde jedoch zur Anpassung an *N. punctiforme* in einigen Punkten abgewandelt (Cohen *et al.* 1994). Die Vorgehensweise war folgende. Die Zellen einer Flüssigkultur von *N. punctiforme* in der exponentiellen Wachstumsphase wurden mittels Zentrifugation bei 1.000 g geerntet und in frischem AA/4, das mit 5 mM MOPS und 2,5 mM NH₄Cl ergänzt war, aufgenommen. Die Zelldichte wurde anhand der Menge an Chlorophyll a bestimmt und durch Verdünnen bzw. weiteres Aufkonzentrieren auf einen Wert von etwa 75 bis 100 µg/ml eingestellt, was erfahrungsgemäß einer Zellzahl von ca. $0,75$ bis 1×10^8 pro ml entspricht. Die beiden Plasmid-enthaltenden *E. coli*-Stämme wurden in LB-Medium bis zur mittleren logarithmischen Phase angezogen, wobei das Medium in beiden Fällen entsprechend der Plasmid-vermittelten Antibiotika-Resistenzen mit 25 µg/ml Chloramphenicol und 30 µg/ml Kanamycin versetzt waren. Die Zelldichten wurden anhand der Absorption bei einer Wellenlänge von 600 nm (OD₆₀₀) bestimmt. Volumina der beiden Kulturen wurden so miteinander gemischt, dass sie eine gleiche Zahl an OD₆₀₀-Einheiten hatten, um Mengenäquivalenz der beiden *E. coli*-Stämme zu gewährleisten. Zur Entfernung der Antibiotika wurde die Mischung einmal mit frischem LB-Medium gewaschen und anschließend auf eine Zelldichte von ca. 9 bis 10 OD₆₀₀-Einheiten eingestellt. Alle Zentrifugationen der *E. coli*-Kulturen erfolgten bei 2.000 g. Pro Konjugationsansatz wurden 0,5 ml der *N. punctiforme*-Suspension mit 0,5 ml der gemischten *E. coli*-Kulturen vermengt, kurz gemischt und in einer Tischzentrifuge bei maximaler Drehzahl für 30 Sekunden pelletiert. Etwa 800 µl des Überstandes wurden mit einer Pipette abgenommen, das Zellpellet durch Pipettieren im Restvolumen resuspendiert und dann auf einen HATF-Filter, der auf eine Platte mit AA-Medium gelegt war, ausgestrichen, wobei das AA-Medium zusätzlich 5 mM MOPS, 2,5 mM NH₄Cl und 0,5 % (v/v) LB-Medium enthielt. Die Platten wurden mit Parafilm verschlossen und über Nacht bei geringer Lichtstärke (ca. 2 W/m²) bei 20° C inkubiert. Am Folgetag wurden die HATF-Filter auf frische AA-Platten mit 5 mM MOPS und 2,5 mM NH₄Cl, jedoch ohne Zusatz von LB-Medium transferiert und für einen weiteren Tag bei geringer Lichtstärke und 20° C inkubiert. Danach wurde die Lichtstärke auf ca. 5 W/m² erhöht und für weitere zwei bis drei Tage inkubiert. Zur Selektion erfolgreicher Exkonjuganten-Kolonien wurden die Filter schließlich auf Platten mit AA-Medium transferiert, die entsprechend der eingebrachten Resistenz-

Kassette das Antibiotikum Neomycin in einer Endkonzentration von 25 µg/ml enthielten. *N. punctiforme*-Kolonien, die in Anwesenheit von Neomycin wuchsen, wurden abschließend auf AA-Platten mit einem Saccharose-Anteil von 5 % (w/v) ausgestrichen, um anhand des auf dem Plasmid pRL271 befindlichen *sacB*-Gens auf doppelte Rekombinationsereignisse zu testen.

2.2.2 Molekularbiologische und biochemische Methoden

2.2.2.1 Isolierung genomischer DNA aus *N. punctiforme*

Zur Isolierung genomischer DNA aus *N. punctiforme* wurden Zellen aus Flüssigkultur mittels 5-minütiger Zentrifugation bei 4 °C mit 1.000 g geerntet und das resultierende Zellpellet zweimal mit 5 M NaCl-Lösung gewaschen. Nach Resuspendierung der Zellen in 1 ml TE-Puffer wurde 1 ml einer frisch hergestellten Lysozym-Lösung (20 mg Lysozym/ml in TE-Puffer) zugegeben, vorsichtig gemischt und 1 h bei 37° C unter gelegentlichem Mischen inkubiert. Anschließend wurden 0,5 ml einer 0,5 M EDTA-Lösung, 1 ml einer frisch hergestellten Proteinase K-Lösung (2 mg Proteinase K/ml in TE-Puffer) und 100 µl SDS-Lösung (20%) zugegeben. Die Lösung wurde gemischt und wiederum 1 h bei 37° C inkubiert. Nach Zugabe von 600 µl 5M NaCl-Lösung und 450 µl CTAB-Lösung (10% (w/v) CTAB und 0,7 M NaCl) wurde 10 min bei 65 °C inkubiert. Danach wurde bei Raumtemperatur 5 min mit 13.000 g zentrifugiert und der Überstand mit einer Pipette abgenommen und in ein frisches Reaktionsgefäß überführt. Die Lösung wurde zweimal mit einem gleichen Volumen Chloroform extrahiert. Aus der wässrigen Phase wurde die DNA durch Zusatz eines doppelten Volumens an 95 %igem Ethanol präzipitiert und durch Zentrifugation bei 13.000 g für 5 min pelletiert. Das DNA-Pellet wurde in 500 µl TE-Puffer aufgenommen und solange mit gleichen Volumina einer Phenol-Chloroform-Mischung (Volumenverhältnis 1:1) extrahiert, bis der Überstand klar blieb. Abschließend wurde die DNA aus der wässrigen Lösung durch Zugabe von 1/10 Volumen 3 M Natriumacetat-Lösung und eines doppelten Volumens an 95%-igem Ethanol erneut präzipitiert und das erhaltene Pellet einmal mit 70%-igem Ethanol gewaschen.

2.2.2.2 Klonierungen

2.2.2.2.1 Allgemeine Methodik

Bei allen Schritten von Klonierungen wie Restriktionsverdau, Ligationen, Plasmidisolierung und Transformation von Bakterien wurden molekularbiologische Standardmethoden eingesetzt (Sambrook und Russell 2001). Die Transformation von Plasmiden und deren Propagation erfolgte in *E. coli*-Zellen, die eine chemische Kompetenz zur Aufnahme von DNA besaßen (siehe Abschnitt 2.2.1.2).

2.2.2.2.2 pRL271-Konstrukt zur Herstellung einer *N. punctiforme*-Mutante

Die Herstellung von Knockout-Mutanten von *N. punctiforme* folgte der Strategie, den Leserahmen des Zielgens durch Einbau einer Antibiotika-Resistenzkassette zu unterbrechen. Aus genomischer DNA von *N. punctiforme* wurde mittels der Polymerase-Kettenreaktion (PCR) ein etwa 4 kbp langes DNA-Fragment,

welches die Sequenz des Zielgens NpF3173 enthielt, amplifiziert. Als PCR-Primer wurden NpF3173KO-FWD und NpF3173KO-REV verwendet, mit denen zugleich eine *SacI*- und eine *PaeI*-Restriktionsschnittstelle in die amplifizierte DNA eingeführt wurden. Die Zusammensetzung der PCR-Ansätze richtete sich nach den Empfehlungen des Herstellers. Das PCR-Programm war aus folgenden Schritten aufgebaut: 1 min Erhitzen bei 94 °C, danach 10 Zyklen mit 10 Sekunden 94 °C zur Strangtrennung, 1 min *Annealing* bei 57 °C und 4 min Strangverlängerung bei 68 °C, danach 24 Zyklen mit 10 Sekunden 94 °C, 1 min *Annealing* bei 59 °C und 4 min und 10 Sekunden bei 68 °C. Die erhaltenen DNA-Amplifikate wurden mittels TA-Klonierung entsprechend der Anleitung des Herstellers in den Vektor pGEM-T einkloniert, wodurch der Vektor pGEM-Npun3173 entstand. Dieser wurde mit dem ohne Überhänge schneidenden Restriktionsenzym *Bsp68I* verdaut. Das geschnittene Konstrukt wurde mit alkalischer Phosphatase dephosphoryliert und die Kanamycin-Resistenzkassette, die aus dem Plasmid pUC4K durch Verdau mit *PvuII* ohne Überhänge ausgeschnitten worden war, einligiert, wodurch das Konstrukt pGEM-Npun3173KAN entstand. Aus diesem wurde das DNA-Stück mit der eingefügten Kanamycin-Resistenzkassette durch einen Doppelverdau mit den Restriktionsenzymen *SacI* und *PaeI* ausgeschnitten und mit dem ebenfalls *SacI*/*PaeI*-doppelverdauten Plasmid pRL271 ligiert, womit das Konstrukt pRL271-Npun3173KAN gebildet wurde. Dieses wurde mit Hilfe eines *triparental mating* konjugativ in die *N. punctiforme*-Zellen eingeschleust (siehe Abschnitt 2.2.1.3).

2.2.2.2.3 Klonierung der Expressionsvektoren für heterologe Expression

Die vollständigen Leserahmen der interessierenden Gene wurden mittels PCR aus der genomischen DNA von *N. punctiforme* amplifiziert, wobei für das NRPS-Modul NpF3173 die Primer NpF3173Exp-FWD und NpF3173Exp-REV zum Einsatz kamen, für das Regulator-Protein NpF3179 die Primer 636/8-FWD und 636/8-REV. In der PCR wurde eine Mischung aus Pfu-Polymerase mit *proofreading*-Aktivität und Taq-Polymerase verwendet. Das PCR-Programm bestand aus einer initialen Denaturierung bei 94 °C für 3 min und 35 Zyklen mit jeweils 30 Sekunden Denaturierung bei 94 °C, 30 Sekunden *annealing* bei 53 °C und 90 Sekunden Elongation bei 72 °C. Zum Abschluss folgte eine 10-minütige Inkubation bei 72 °C. Die erhaltenen DNA-Fragmente wurden gemäß der Anweisungen des Herstellers direkt in den Expressionsvektor pBAD-TOPO ligiert, der eine Expression der Proteine mit C-terminalem 6-fach His-Tag erlaubt.

2.2.2.2.4 Klonierung der DNA-Fragmente für den *Electrophoretic mobility shift assay*

Zur Vorbereitung des EMSA wurden aus der genomischen DNA von *N. punctiforme* zwei DNA-Fragmente mit einer Länge von ca. 140 bp amplifiziert, von denen das als DNA(BIND) bezeichnete die vermeintliche Bindungsstelle des Regulators enthielt, während DNA(NONBIND) zufällig aus der Sequenz des betreffenden Genclusters ausgewählt war und die Bindestelle nicht enthielt. Zur Amplifizierung von DNA(BIND) diente das Primer-Paar Reg-Bind-FWD und Ref-Bind-REV, zur Amplifizierung von DNA(NONBIND) wurden die Primer NonReg-Bind-FWD und NonReg-Bind-REV benutzt. Die PCR wurde mit folgendem Programm durchgeführt: Es wurde initial 3 min bei 94 °C denaturiert, woran sich 35 Zyklen mit jeweils 30 s Denaturierung bei 94 °C, 30 s *annealing* bei 52 °C und 1 min Elongation bei 72 °C

anschlüssen. Abschließend wurde 10 min bei 72 °C inkubiert. Die entstandenen DNA-Amplifikate wurden direkt gemäß Herstellerangaben in den Vektor pGEM-T einkloniert.

2.2.2.3 *Reinigung von Proteinen mit His-Tag*

Die Aufreinigung der Proteine, die mit Hilfe des pBAD-TOPO-Vektors exprimiert wurden, erfolgte unter Ausnutzung des C-terminalen His-Tags unter nativen Bedingungen. Die Zellpellets der Expressionskulturen von *E. coli* wurden in Lysis-Puffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM Imidazol, 0,5 % (v/v) Tween 20, pH 8,0) resuspendiert, wobei das Volumenverhältnis von Bakterienkultur zu resuspendiertem Pellet 50:1 betrug. Alle weiteren Angaben beziehen sich auf ein Volumen an resuspendierten Zellen von ca. 5 ml. Die Zellsuspension wurde mit Lysozym zu einer Endkonzentration von 1 mg/ml versetzt und 30 min auf Eis inkubiert. Anschließend wurden die Zellen auf Eis gelagert mittels eines Ultraschall-Stabs durch sechs jeweils etwa 10 s andauernde Impuls von etwa 200 W aufgeschlossen und das Lysat 30 min bei 4 °C mit 10.000 g zentrifugiert. Der klare Lysat-Überstand wurde mit der aufgeschwemmten Nicke-Agarose versetzt, wobei das Volumenverhältnis von Lysat-Überstand zu Nickel-Agarose 4:1 betrug, und 60 min bei 4 °C leicht geschüttelt. Alle weiteren Arbeitsschritte erfolgten ebenfalls bei 4 °C. Die Lysat/Nickel-Agarose-Mischung wurde in eine 5 ml-Plastiksäule eingefüllt und über den Auslass soweit entleert, bis sich die Agarose-Matrix vollständig abgesetzt hatte. Die Säule wurde dann dreimal mit je 4 ml Waschpuffer (50 mM NaH₂PO₄, 350 mM NaCl, 30 mM Imidazol, 0,5 % (v/v) Tween 20, pH 8,0) gewaschen und schließlich mit Elutionspuffer (50 mM NaH₂PO₄, 350 mM NaCl, 250 mM Imidazol, 0,5 % (v/v) Tween 20, pH 8,0) eluiert, wobei einzelne Fraktionen von etwa 500 µl gesammelt wurden. Fraktionen, die das exprimierte Protein enthielten, wurden durch SDS-PAGE und anschließende Färbung der Gele (siehe Abschnitt 2.2.2.5) identifiziert. Diese Fraktionen wurden vereinigt und gegen Puffer mit 20 mM Tris (pH 8) und 150 mM NaCl bei 4 °C über Nacht dialysiert. Zur längeren Aufbewahrung wurde die Proteinlösung mit Glycerol versetzt (Endkonzentration 30 % (v/v)) und aliquotiert bei -20 °C gelagert.

2.2.2.4 *Konzentrationsbestimmung von Protein-Lösungen*

Die Bestimmung der Konzentration von Protein-Lösungen erfolgte mit dem „Bio-Rad Protein Assay“, der auf der Methode nach Bradford beruht (Bradford 1976). Bei der Anwendung wurde den Anweisungen des Herstellers gefolgt.

2.2.2.5 *Protein-Gelelektrophorese und Western Blot*

Die elektrophoretische Auftrennung von Proteinen erfolgte in Polyacrylamid-Gelen mit einem diskontinuierlichen System, bestehend aus Sammel- und Trenngel (Laemmli 1970) unter Zusatz von SDS, also unter denaturierenden Bedingungen (SDS-PAGE). Für die Durchführung wurde das „Mini Protean II“-System von Bio-Rad benutzt, die verwendeten Materialien und Puffer entsprachen der Standardmethodik (Eckert und Kartenbeck 1997). Die Acrylamid-Stammlösung wies eine Gesamtkonzentration von T=30% auf und einen Vernetzungsgrad von C=2,7%. Zur Analyse der Expression des Regulator-Proteins NpF3179

wurde ein Trenngel mit 15 % Acrylamid-Anteil benutzt, zur Analyse des NRPS-Moduls NpF3173 ein Trenngel mit 7 % Acrylamid-Anteil. In beiden Fällen besaßen die Sammelgele einen Acrylamid-Anteil von 4 %. Alle Gelläufe erfolgten bei einer konstanten Stromstärke von 30 mA /Gel. Nach dem Gellauf wurden die Gele entweder mittels einer kolloidalen „Coomassie Brilliantblau G-250“-Lösung (Pierce) angefärbt oder für immunchemische Nachweise auf Membrane transferiert. Zur Färbung wurden die Gele kurz mit destilliertem Wasser gespült und dann über Nacht bei Raumtemperatur in der Färbelösung inkubiert. Diese wurde anschließend abgegossen und überschüssiger Farbstoff durch Spülen mit Wasser entfernt.

Für immunchemische Nachweise wurden die Proteine mittels *Western Blotting* aus dem Gel auf PVDF-Membrane transferiert. Dies geschah als so genanntes „Tank Blotting“ unter Verwendung der zum „Mini Protean II“ von Bio-Rad gehörenden *Blotting*-Apparatur. Der *Blotting*-Stapel wurde gemäß der Angaben des Herstellers zusammengebaut, wobei unterhalb und oberhalb von Gel bzw. Membran jeweils zwei Lagen Filterpapier (3MM von Whatman) geschichtet wurden. Der Transfer fand in Standard-Puffer (25 mM Tris, 192 mM Glycin) ohne Methanol bei 4 °C für 1 h bei einer konstanter Spannung von 100 V (entsprechend einer anfänglichen Stromstärke von ca. 200 mA und einer Endstromstärke von ca. 400 mA) statt. Nach dem Transfer wurde die Membran durch Schwenken in TBS (Tris-gepufferte Salzlösung: 10 mM Tris, 150 mM NaCl, pH 7,5), der 5 % Magermilchpulver und 0,5 % Tween 20 zugesetzt waren, über Nacht bei 4 °C behandelt, um unspezifische Bindungsstellen zu blockieren. Nach dreimaligem Waschen mit TBST (TBS versetzt mit 0,5 % Tween 20) wurde mit dem Primärantikörper 1 h bei Raumtemperatur inkubiert, wobei der Antikörper in TBST mit 3 % Magermilchpulver gelöst war. Anschließend wurde wiederum dreimal mit TBST gewaschen. Es folgte die Inkubation mit dem sekundären, Alkalische-Phosphatase-gekoppelten Antikörper, ebenfalls 1 h bei Raumtemperatur. Nach dreimaligem Waschen mit TBS wurde die Antikörperbindung colorimetrisch über die Umsetzung der Substrate 5-Brom-4-Chlor-3-Indolyl-Phosphat (BCIP) und Nitroblau-Tetrazolium (NBT) nachgewiesen (Eckert und Kartenbeck 1997).

2.2.2.6 Electrophoretic mobility shift assay (EMSA)

Die zu testenden DNA-Fragmente wurden durch Restriktionsverdau mit *Bst*ZI aus dem pGEM-T-Vektor (vgl. Abschnitt 2.2.2.2.4) ausgeschnitten und mit Hilfe des „Qiaquick PCR Purification Kit“ gemäß Herstellerangaben gereinigt. Die DNA-Fragmente wurden am 5'-Ende radioaktiv mit ³²P markiert. Dazu wurden jeweils ca. 400 ng DNA in einem 50 µl-Ansatz mit 10 µl PEG-6000, 5 µl [γ -³²P]-ATP (110 Tbq/mmol) und T4-Polynucleotidkinase 30 min bei 37 °C inkubiert. Der Reaktionsansatz wurde anschließend zur Abtrennung der radioaktiven Nucleotide über eine kurze Gelfiltrationssäule zentrifugiert und die Volumenaktivität des die markierte DNA enthaltenden Eluats bestimmt.

Für den Bindungstest wurden je Ansatz 100 ng des Regulator-Proteins NpF3179 und eine Menge an DNA, die etwa 10.000 cpm aufwies, kombiniert. Als Kontrollen wurden darüber hinaus auch Ansätze ohne Protein, ohne DNA und mit verschiedenen Mengen an unmarkierter DNA als Kompetitor untersucht. Das Protein wurde zunächst in einem 30 µl-Ansatz in Bindungspuffer (10 mM Tris/HCl pH 8, 10 mM KCl, 10 mM MgCl₂, 0,1 µg/ml BSA, 1 mM DTT, 12 % (v/v) Glycerol) 10 min bei Raumtemperatur vorinkubiert und nach Zugabe der markierten DNA 30 min ebenfalls bei Raumtemperatur weiter inkubiert. Bei Konkurrenzexperimenten mit unmarkierter DNA wurde zunächst das Protein allein 5 min vorinkubiert, die unmarkierte DNA zugegeben und für weitere 5 min inkubiert, bevor die Zugabe der markierten DNA und

die 30-minütige Hauptinkubation erfolgten. Die Reaktionsansätze wurden schließlich mittels Gelelektrophorese in Polyacrylamid-Gelen aufgetrennt, um Änderungen in der elektrophoretischen Beweglichkeit nach Komplexbildung analysieren zu können.

Die elektrophoretische Auftrennung wurde mit einem „Protean II xi“-System der Firma Bio-Rad durchgeführt. Zur Herstellung von 40 ml Gellösung zum Gießen von Gelen mit 7,5 % Acrylamid-Gehalt wurden 10 ml 30 %ige Acrylamid-Stammlösung (Verhältnis Acrylamid:Bis-Acrylamid=30:0,8) mit 2 ml 10x TBE-Laufpuffer (108 g Tris-Base, 55 g Borsäure und 20 ml 0,5 M EDTA-Lösung/pH 8,0 in 1000 ml), 28 ml deionisiertem Wasser, 500 µl 10 %ige (w/v) APS-Lösung und 50 µl TEMED gemischt. Der Gellauf erfolgte mit 0,5x TBE als Laufpuffer bei 4 °C. Die Gele liefen zunächst leer bei konstant 200 V, bis eine Stromstärke von 15 mA pro Gel erreicht war. Danach wurden die Bindungsreaktionsansätze auf die Gele geladen und der Lauf bei 15 mA je Gel für vier bis sechs Stunden fortgesetzt. Die Gele wurden nach dem Lauf auf 3MM-Filterpapier (Whatman) transferiert und getrocknet. Zur Sichtbarmachung der radioaktiven Signale wurden PhosphorImager-Schirme mit den getrockneten Gelen exponiert. Die Auswertung der Signale erfolgte mit der „Quantity One“-Software von Bio-Rad.

2.2.2.7 ATP-Pyrophosphat-Austausch-Assay

Zur Überprüfung der Adenylierungsaktivität des NRPS-Moduls NpF3173 wurde die ATP-Pyrophosphat-Austausch-Reaktion eingesetzt (Gevers, Kleinkauf und Lipmann 1969). Die Reaktionsansätze enthielten in einem Endvolumen von 100 µl 50 mM Tris, pH 8, 100 mM Magnesiumacetat, 2 mM ATP, 2 mM DTT, 0,2 mM EDTA, 0,5 mM der zu testenden Aminosäure und ca. 25 µg des Enzyms. Die Reaktion wurde durch Zugabe von Tetranatrium-Pyrophosphat (Endkonzentration: 0,2 mM) und 0,15 µCi Tetranatrium-[³²P]-Pyrophosphat gestartet und 20 min bei 28 °C im Wasserbad inkubiert. Die Umsetzung wurde durch Zugabe von 0,5 ml Stopp-Lösung (1,5 % (w/v) Aktivkohle, 100 mM Tetranatrium-Pyrophosphat und 0,4 M Perchlorsäure) abgestoppt. Die Reaktionsansätze wurden auf Papierfilter abgesaugt, dreimal mit je 5 ml Wasser gewaschen, getrocknet und nach Zugabe von 3 ml Flüssig-Szintillationsmischung im Szintillationszähler ausgewertet.

2.2.2.8 Probengewinnung aus *N. punctiforme*-Kulturen

Sowohl die Extrakte aus Zellpellets von *N. punctiforme* als auch Proben aus den Mediumüberständen von Flüssigkulturen wurden mit Hilfe von Festphasenextraktions (SPE)-Kartuschen (Sep-Pak Plus C18 von Waters) hergestellt. Zur Gewinnung von Extrakten aus Zellpellets wurden die Zellen mittels Zentrifugation geerntet, das Zellpellet einmal mit frischem AA/4-Medium gewaschen und schließlich in entionisiertem Wasser resuspendiert. Bezogen auf eine dicht gewachsene Flüssigkultur von 200 ml betrug das Volumen des resuspendierten Pellets etwa 10 ml. Die Zellen wurden bei 4 °C in einer Retsch-Kugelmühle aufgeschlossen. Dazu wurde die Zellsuspension in Eppendorf-Reaktionsgefäße überführt, mit einem gleichem Volumen einer Glaskugelmischung (50% 0,5 mm und 50% 0,2 mm Durchmesser) versetzt und auf höchster Stufe 1 min lang geschüttelt. Direkt danach wurden die Proben in flüssigem Stickstoff eingefroren, aufgetaut und wiederum 1 min geschüttelt. Einfrieren, Auftauen und Schütteln wurden ein

weiteres Mal wiederholt. Nachfolgend wurde 15 min bei 4 °C mit 13.000 g zur Entfernung der Zelltrümmer und Glaskugeln zentrifugiert. Der Überstand wurde abgenommen und gesammelt. Die SPE-Kartuschen wurden gemäß Herstellerangaben mit 2 ml Methanol und Waschen mit 2 ml Reinwasser vorbereitet. Der gesammelte Zentrifugationsüberstand wurde mit einer Spritze durch die SPE-Kartusche gedrückt. Diese wurde mit 2 ml wässriger 5 %iger (v/v) Methanol-Lösung gespült und mit 2 ml Methanol eluiert. Das Eluat wurde in einer Vakuum-Zentrifuge vollständig eingedampft und der Rückstand anschließend in einer Methanol/Wasser-Mischung (Volumenverhältnis 3:2) gelöst, wobei das Volumen so gewählt wurde, dass der Konzentrierungsfaktor in Bezug auf das Ausgangsvolumen an Zellkultur bei etwa 1.000 lag. Mittels eines 0,45 µm-Spritzenvorsatzfilters wurden die Proben von nicht gelösten Partikeln befreit, bevor sie in die HPLC-Anlage eingespritzt wurden.

Zur Gewinnung von Proben des Kulturmediums wurden die Überstände der Zellernte gegebenenfalls bei hoher relativer Zentrifugalkraft zentrifugiert, um sämtliche Zellreste zu entfernen. Die klaren Überstände wurden, wie für die Zellpellet-Proben beschrieben, mit SPE-Kartuschen weiter konzentriert.

2.2.2.9 HPCL-Analytik

Die HPLC-Trennungen der Zellextrakt- und Kulturüberstandsproben von *N. punctiforme* wurden auf einer HPLC-Anlage von Shimadzu durchgeführt. Die Anlage bestand aus dem Systemkontroll-Element SCL-10AVP, einer Pumpe LC-10Ai, dem automatischen Probengeber SIL-10A und dem Fraktionssammler FRC-10A. Als Detektionssystem diente ein Photodioden-Array-Detektor (PDA-Detektor) des Typs SPD-M-10AVP. Die chromatographischen Trennungen erfolgten auf einer „SymmetryShield RP18“-Säule von Waters mit einer Partikelgröße von 3,5 µm, einem Innendurchmesser von 4,6 mm und einer Länge von 100 mm, der eine mit demselben Sorbens gefüllte Vorsäule der Dimension 3,9 mm x 20 mm vorgeschaltet war. Als Eluenten dienten Wasser (Eluent A) und Acetonitril (Eluent B). Die HPLC-Läufe fanden bei einer Flussrate von 1 ml/min mit einem Wasser-Acetonitril-Gradienten statt. Sowohl das Wasser als auch das Acetonitril waren mit 0,05 % Trifluoressigsäure als Ionenpaar-Reagenz versetzt. Das Gradientenprogramm war folgendermaßen gestaltet: Zu Beginn des Laufes betrug der Anteil der Eluenten A: 80 % und B:20 %; beginnend nach 1 min verlief der Gradient linear über 20 min bis zum Erreichen von A: 30 % und B: 70 %; danach stieg der Anteil von B innerhalb von 1 min bis 100 % an und wurde schließlich innerhalb von 3 min auf den ursprünglichen Wert von 20 % abgesenkt.

2.2.3 Bioinformatische Methoden

2.2.3.1 Datengewinnung

Die Aminosäuresequenzen der bakteriellen PKS des Typs I wurden mit Hilfe der mikrobiellen Genom-Plattform des NCBI unter Benützung der dort zur Verfügung gestellten BlastP-Suchfunktion (Altschul *et al.* 1990; Altschul *et al.* 1997) gewonnen. Der Erwartungswert war dabei der Grundeinstellung entsprechend auf 10 gesetzt. Als Suchsequenz diente die Aminosäure-Sequenz von DEBS1, des ersten PKS-Proteins der Erythromycin-Biosynthese, aus *Saccharopolyspora erythraea*. Gesucht wurde gegen 138 vollständige

eubakterielle Genome, 20 vollständige archaebakterielle Genome und drei zum Zeitpunkt der Untersuchung (Februar 2005) noch in Bearbeitung befindliche Genome der Cyanobakterien *Anabaena variabilis*, *Crocospaera watsonii* und *Nostoc punctiforme*. Die drei letztgenannten Genome wurden mit einbezogen, um den Datenbestand an cyanobakteriellen Genomen zu vergrößern, da bekannt ist, dass Cyanobakterien viele Sekundärmetabolite dieses Typs enthalten (Bode und Müller 2005). Die Ergebnisse der BlastP-Suche wurden manuell überprüft, damit offensichtlich unpassende Ergebnisse nicht in den Datensatz gelangten.

Um die Aminosäure-Sequenzen von FAS zu gewinnen, wurde ebenfalls die NCBI-Genom-Plattform benutzt. In diesem Fall wurde sowohl mit der vollständigen Sequenz der FAS aus *Rattus norvegicus* als auch mit den einzelnen Sequenzen der bakteriellen FAS-Komponenten (FabF, FabH, FabD, FabG) aus *Escherichia coli* als Suchsequenzen gesucht. Alle BlastP-Ergebnisse wurden manuell auf ihre Stimmigkeit überprüft.

2.2.3.2 Domänen-Analyse

Die Aminosäure-Sequenzen der modularen PKS-I wurden mit Hilfe des Internet-basierten Programms SEARCHPKS (Yadav, Gokhale und Mohanty 2003) analysiert. Dieses Programm ermöglichte die Visualisierung der in den Proteinen enthaltenen Module und Domänen, die Zuweisung von Substratspezifitäten von AT-Domänen und das automatische Ausschneiden von Domänen-Sequenzen anhand konservierter Domänen-Grenzen. Da die Zuweisung von Substratspezifitäten nicht in allen Fällen mit ausreichender Sicherheit automatisch erfolgen konnte, wurden die AT-Domänen in vier Gruppen eingeteilt: (1) AT-Domänen, deren Spezifität aufgrund biochemischer Untersuchungen bereits experimentell charakterisiert worden war; (2) AT-Domänen, deren Spezifitätsangaben ausschließlich auf der Voraussage des SEARCHPKS-Programms beruhte; (3) AT-Domänen, deren Spezifität manuell durch Analyse der an der Substraterkennung beteiligten Aminosäure-Reste zugewiesen wurde; (4) AT-Domänen mit unklarer Substratspezifität.

Im Fall der modularen PKS-I aus *Streptomyces avermitilis* wurden die Domänen-Informationen direkt von der Website des *S. avermitilis*-Genomprojekts (<http://avermitilis.ls.kitasato-u.ac.jp>) bezogen.

2.2.3.3 Sequenz-Alignments

Die Aminosäure-Sequenz-Alignments wurden mit den Programmen ClustalW und ClustalX (Thompson, Higgins und Gibson 1994) erstellt. Als Gewichtungsmatrices dienten die Gonnet-Matrix (Gonnet, Cohen und Benner 1992) oder die BLOSUM 62-Matrix (Henikoff und Henikoff 1992). Die Alignment-Parameter waren anfänglich für das *gap opening* auf 10, für das *gap extention* auf 0,2 gesetzt und wurden gegebenenfalls variiert. Die Alignments wurden manuell unter Zuhilfenahme des Programms MacClade (Maddison und Maddison 2000) editiert. Bereiche mit großen Insertionen oder Deletionen wurden aus den Alignments entfernt, ebenso Bereiche mit sehr großer Sequenz-Variabilität, die auf vollständige Mutationsabsättigung schließen ließen. Zur Verbesserung der Alignments von AT- und KS-Domänen wurden die Sekundärstrukturen ausgewählter Domänen mittels des PSIPRED-Servers

(McGuffin, Bryson und Jones 2000) vorhergesagt und mit den bekannten Kristallstrukturen der *E. coli*-Proteine FabD (Serre *et al.* 1995) bzw. FabF (Moche *et al.* 1999) verglichen, um die korrekte Positionierung konservierter Sekundärstruktur-Elemente zu gewährleisten. Zur Erstellung von Nucleotid-Alignments codierender Sequenzabschnitte wurden zunächst die entsprechenden Aminosäure-Alignments erstellt und diese mit Hilfe des Programms DAMBE (Xia und Xie 2001) unter Berücksichtigung der Codon-Grenzen in die Nucleotid-Alignments umgewandelt.

2.2.3.4 Phylogenetische Analysen

Zur Erstellung der phylogenetischen Rekonstruktionen wurden verschiedene Verfahren eingesetzt: das Bayes-Verfahren, die Distanz-basierte *Neighbour joining*-Methode, der *Maximum Parsimony*-Ansatz und *Maximum Likelihood*-Berechnungen, gestützt auf den *Quartet puzzling*-Ansatz. Für eine ausführliche und vergleichende Darstellung dieser Methoden wird auf aktuelle Übersichtsartikel (Lio und Goldman 1998; Holder und Lewis 2003) und drei Gesamtdarstellungen in Buchform (Salemi und andere 2003; Felsenstein 2004; Nielsen und andere 2005) verwiesen.

2.2.3.4.1 Bayes-Verfahren

Zur Berechnung der phylogenetischen Rekonstruktionen nach dem Bayes-Verfahren wurde das Programm „MrBayes“ in der Version 3 benutzt (Ronquist und Huelsenbeck 2003). Die am besten geeigneten Modelle der Proteinevolution wurden mit der in „MrBayes“ integrierten Modelltest-Funktion, bei der alle implementierten Modelle durchlaufen werden, und mit dem Programm „Prottest“ (Drummond und Strimmer 2001; Guindon und Gascuel 2003; Abascal, Zardoya und Posada 2005) ermittelt. Für die Analyse wurde aufgrund der erhaltenen Ergebnisse dasjenige Modell ausgewählt, welches den statistischen Kriterien am besten entsprach und in allen Phylogenie-Programmen, die parallel auf einen Datensatz angewendet werden sollten, implementiert waren. Diesen Kriterien entsprach in allen Fällen das JTT-Modell der Aminosäure-Substitution (Jones, Taylor und Thornton 1992) unter Berücksichtigung einer Gamma-Verteilung zur Modellierung der unterschiedlichen Austauschhäufigkeit an verschiedenen Sequenz-Positionen (JTT+ γ). Die kontinuierliche Gamma-Funktion wurde durch eine diskrete Gamma-Verteilung mit vier Kategorien angenähert. Die Metropolis-gekoppelte Markov-Ketten-Monte-Carlo-Analyse (MCMC) bestand aus vier unabhängigen Ketten und lief solange, bis aufgrund der berechneten *Likelihood*-Werte eine stabile Konvergenz angenommen werden konnte. Die Markov-Kette wurde dabei in jeder hundersten Generation abgefragt. Stabile Konvergenz wurde anhand von Plots der *Likelihood*-Werte und den Standardabweichungen zwischen zwei unabhängigen Berechnungen beurteilt. Von einer Konvergenz war auszugehen, wenn die so genannten Akzeptanzraten innerhalb der „kalten“ Kette im Bereich von 10 bis 70 % lagen und außerdem auch die Werte der Akzeptanzraten für den Wechsel zwischen den vier unabhängigen Ketten Werte von 10 bis 70 % aufwiesen (siehe das Handbuch zum MrBayes“-Programm (Ronquist, Huelsenbeck und van der Mark 2005)). Alle vor dem Erreichen des so festgelegten Konvergenz-Status abgefragten Bäume wurden als *Burn in*-Phase nicht weiter berücksichtigt. Die nach Erreichen der Konvergenz berechneten Bäume dienten gemäß der *Majority*-Regel zur Kalkulation der Consensus-Bäume und der posterioren *Clade probability*-Werte.

Im Einzelnen wurden für die Rekonstruktion der Phylogenie der KS-Domänen bzw. der KS-Proteine von PKS und FAS (Abschnitt 3.1.1) 1,5 Millionen Generationen berechnet, für die Analyse der PKS-III (Abschnitt 3.1.2) ebenfalls 1,5 Millionen Generationen und für Rekonstruktion der Phylogenien der AT- und KS-Domänen aus bakteriellen PKS-I 3.2.3) jeweils 4 Millionen Generationen. Der phylogenetische Baum der KS-Domänen aus ausgewählten PKS-I-Clustern von Streptomycceten basiert auf einer MCMC-Analyse über 10 Millionen Generationen

Für die phylogenetische Analyse der verschiedenen Domänen-Typen der PKS-I aus *S. avermitilis* wurde ein kombinierter Datensatz verwendet, der aus den Aminosäure-Sequenzen und den entsprechenden Nucleotid-Sequenzen bestand. Dabei wurde auf die Aminosäure-Daten das JTT+ γ -Modell angewandt, auf die Nucleotid-Daten das „allgemeine Zeit-umkehrbare Modell“ (englisch: *general time reversible model*) (Yang, Goldman und Friday 1994), wobei eine Einteilung nach Codon-Positionen erfolgte. Die Berechnungen umfassten in diesen Fällen vier Millionen Generationen bei den KS-Domänen und AT-Domänen, zwei Millionen Generationen bei den DH-Domänen und 2,5 Millionen Generationen bei den KR-Domänen. Für alle vier Domänen-Typen lag die Grenze der Standardabweichungen zur Beurteilung der Konvergenz bei 0,005.

2.2.3.4.2 *Neighbour joining*-Methode

Zur Berechnung von Phylogenien nach der *Neighbour joining*-Methode wurde das PHYLIP-Software-Paket in der Version 3.6 herangezogen (Felsenstein 2002). Im Fall von Protein-Sequenzen wurden zunächst mit dem Modul Seqboot Pseudoreplikate des zu analysierenden Sequenz-Alignments erstellt. Auf dieser Basis wurden im nächsten Schritt mit dem Modul Protdist in entsprechende Distanzmatrices berechnet, wobei das JTT+ γ -Modell benutzt wurde. Der α -Parameter der Gamma-Verteilung wurde mit Hilfe des Programms TREE-PUZZLE (Schmidt *et al.* 2002) ermittelt. Nachfolgend wurden die Distanzmatrices mit dem Modul Neighbor ausgewertet und aus den erhaltenen Bäumen ein Consensus-Baum gemäß der *Majority*-Regel errechnet. Die Zahl der benutzten Pseudoreplikate betrug 1000.

Im Fall der Berechnung einer Spezies-Phylogenie anhand von DNA-Sequenzen der kleinen ribosomalen Untereinheit (Abschnitt 3.2.3.3) wurde auf dieselbe Weise vorgegangen, wobei anstatt des Moduls Protdist das Modul Dnadist zum Einsatz kam. Als Evolutionsmodell diente bei dieser Rekonstruktion das F84-Modell (Kishino und Hasegawa 1989) unter Berücksichtigung von positionsspezifischen Heterogenitäten von Austauschhäufigkeiten.

2.2.3.4.3 *Maximum Parsimony*-Analyse

Berechnungen nach der *Maximum Parsimony*-Methode wurden mit Hilfe des Programms PAUP* 4.0b10 (Swofford 2002) durchgeführt. Konstante und nicht-informative Positionen wurden nicht berücksichtigt, Sequenz-Lücken als „fehlende Daten“ interpretiert. Die Berechnungen erfolgten mit der heuristischen Such-Option bei stufenweiser Addition der Äste und aktiver *tree-bisection reconnection*-Option (TBR) für das Austauschen von Ästen. Aus allen erhaltenen besten Bäumen wurde ein strikter Consensus-Baum errechnet.

2.2.3.4.4 *Maximum Likelihood-Analyse*

Für phylogenetische Rekonstruktionen nach dem *Maximum Likelihood*-Verfahren wurde das Programm TREE-PUZZLE (Schmidt *et al.* 2002) benutzt. Dieses Verfahren wurde aufgrund des hohen Rechenaufwands nur für kleinere Datensätze mit bis zu 60 Sequenzen im Alignment eingesetzt. Als Modell der Protein-Evolution diente wie bei den anderen Methoden das JTT+ γ -Modell, wobei die kontinuierliche Gamma-Funktion durch eine diskrete Verteilung mit vier Kategorien angenähert wurde.

2.2.3.5 *Selektionsanalyse*

Mögliche Selektionsereignisse in den PKS-I-Genen von *S. avermitilis* wurden anhand des Verhältnisses zwischen synonymen und nicht-synonymen Basentauschen untersucht. Für diese Analysen wurde das Programm MEGA (Kumar *et al.* 2001; Kumar, Tamura und Nei 2004) in der Version 3 benutzt. Dabei wurde das Ausmaß an Nucleotid- und Aminosäure-Austauschen über die unkorrigierten Unterschiede (p-Distanzen) abgeschätzt. Dieser Wert hat gegenüber anderen Distanzmaßen den Vorteil, dass er bei großen Datensätzen mit einer relativ geringen Zahl an ausgewerteten Positionen aufgrund seiner kleineren Varianz stabilere Resultate ergibt (Nei 2000). Die Zahl der synonymen (p_S) und der nicht-synonymen (p_N) Basenaustausche je Position wurden mit der modifizierten Nei-Gojobori-Methode (Zhang, Rosenberg und Nei 1998) berechnet. Die p-Distanzen zwischen den Nucleotid-Sequenzen bzw. Aminosäure-Sequenzen wurden bei paarweiser Deletion bestimmt. Die Berechnung der Standardfehler basierte auf der *Bootstrap*-Methode mit 1000 Pseudoreplikaten. Positive Selektion wurde über die Nullhypothese $H_0: p_S=p_N$ gegen die Alternative $H_1: p_S>p_N$ getestet. Hierzu wurden die durchschnittlichen Werte von p_S und p_N sowie deren Varianzen mit Hilfe des Codon-basierten Z-Tests verglichen (Nei 2000).

3 Ergebnisse

3.1 Evolutionäre Stellung der verschiedenen PKS-Typen

3.1.1 Evolutionäre Zusammenhänge zwischen FAS und PKS

Um die evolutionären Beziehungen zwischen den unterschiedlichen Typen von FAS und PKS zu untersuchen, wurde ein Datensatz aus Aminosäure-Sequenzen erstellt, der die KS-Komponenten aller System-Typen außer PKS-III enthielt. Letztere wurden in anderen Datensätzen analysiert (siehe Abschnitt 3.1.2). Proteine bzw. Domänen mit KS-Aktivität wurden gewählt, da diese die längsten Sequenzen aufweisen und nur wenige sehr variable Bereiche haben, so dass sie hinsichtlich ihres Informationsgehalts für eine phylogenetische Analyse besonders gut geeignet sind. Der Datensatz enthielt ausgewählte Vertreter aller Gruppen, wobei insbesondere biochemisch charakterisierte Komponenten in den Datensatz aufgenommen wurden. Aus dem editierten Sequenz-Alignment wurden dann unter Anwendung des Bayes-Verfahrens sowie der *Maximum Parsimony*- und der *Neighbor-Joining*-Methode phylogenetische Bäume rekonstruiert. Das Ergebnis des Bayes-Verfahrens ist in **Abb. 8** gezeigt.

Mit dem MP-Verfahren und der NJ-Methode wurde dieselbe Baum-Topologie erhalten, d. h. alle Cladi und Subcladi ließen sich reproduzieren. Geringfügige Unterschiede traten nur in der Anordnung innerhalb der Subcladi auf (Daten nicht gezeigt). Ausgehend von der aus archaeobakteriellen FabH-Homologen bestehenden Außengruppe, setzen sich die ersten beiden Gruppen aus Vertretern der FAS-II zusammen, wobei die homologen FabF und FabH phylogenetisch aufgetrennt werden. Die FabF-Gruppe enthält bakterielle Sequenzen unter Einschluss der spezialisierten mycobakteriellen FabF-Homologen sowie Sequenzen aus mitochondrial und plastidär lokalisierten Proteinen aus Eukaryoten. Diese werden zwar im Zellkern-Genom codiert, die translatierten Proteine jedoch in die Organellen transloziert. Dabei zweigen erwartungsgemäß die plastidären FabF-Sequenzen direkt von solchen aus Cyanobakterien ab, was ihren cyanobakteriellen Ursprung widerspiegelt, während die mitochondrialen Proteine in der Nähe des FabF aus *Agrobacterium tumefaciens* lokalisiert sind, worin sich ihre Herkunft aus endosymbiontischen α -proteobakteriellen Vorläufern ausdrückt.

Die nächste Gruppe vereint FAS-I-Sequenzen aus Bakterien der CMN-Gruppe (Corynebakterien, Mycobakterien und Nocardien) und Pilzen, die jeweils eine abgeschlossene homogene Untergruppe bilden. Die völlige Übereinstimmung der Domänenanordnung in FAS-I aus Bakterien und Pilzen wurde erstmals mit der Sequenzierung und Analyse der FAS-I aus *Brevibacterium ammoniagenes* gefunden (Meurer *et al.* 1992) und später für die Synthase aus *Mycobacterium tuberculosis* bestätigt (Fernandes und Kolattukudy 1996). Daneben besteht auch eine große Ähnlichkeit in der Aminosäure-Zusammensetzung, die im Fall von *B. ammoniagenes* im Vergleich mit der Hefe *Saccharomyces cerevisiae* bei 30 % bezüglich FAS1 und 46 % bezüglich FAS2 liegt (Meurer *et al.* 1992).

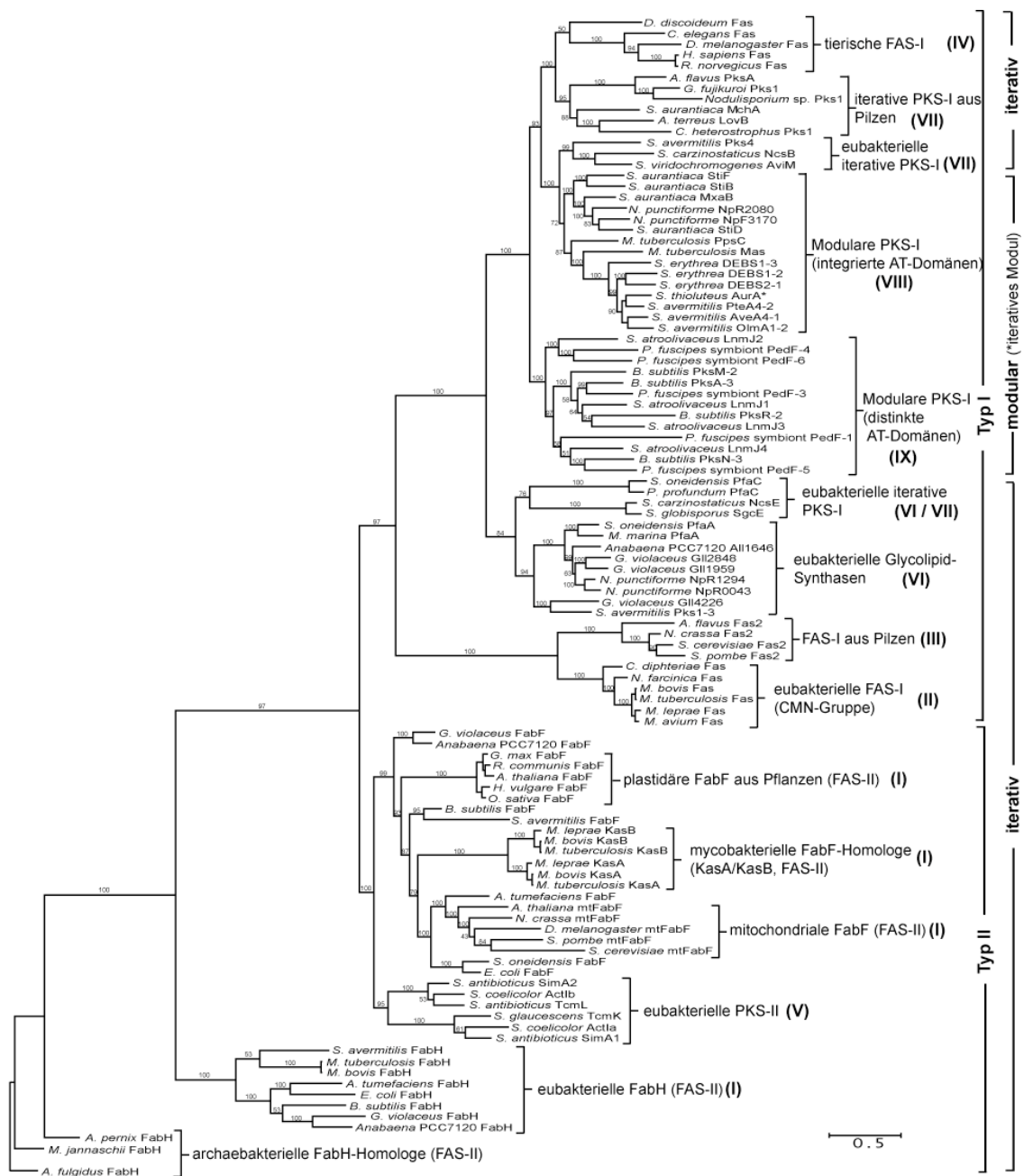


Abb. 8: Phylogenie von KS-Domänen und -Proteinen aus FAS und PKS.

Der Baum wurde mit dem Bayes-Verfahren rekonstruiert. Die Zahlen über den Zweigen geben die *Clade Probability*-Werte an, die Längen der Zweige die Zahl der Aminosäure-Austausche pro Position. Die Bezeichnung der phylogentischen Gruppen und Untergruppen sowie die römischen Zahlen in Klammern beziehen sich auf die Einteilung in **Abb. 4** der Einleitung. Die Angaben am rechten Rand bezeichnen den Strukturtyp und den Funktionsmodus der jeweiligen Systeme.

Die Positionierung als Unterguppen einer gemeinsamen Gruppe, die sich phylogenetisch klar von den anderen Synthasen abtrennen, lässt darauf schließen, dass die genannten Übereinstimmungen auf einen gemeinsamen evolutionären Ursprung zurückzuführen sein könnten.

Im phylogenetischen Baum folgt als nächste Gruppe die der bakteriellen Glycolipid-Synthasen und charakterisierter PKS-I mit iterativem Funktionsmodus. *Photobacterium profundum* und *Shewanella oneidensis* sind marine Bakterien, die über die Fähigkeit verfügen, mehrfach ungesättigte Fettsäuren (PUFA, von englisch: *poly unsaturated fatty acid*) des ω -3-Typs herzustellen wie z.B. Docosaheptaensäure (22:6 ω 3) und Eicosapentaensäure (20:5 ω 3) (Metz *et al.* 2001; Wallis, Watts und Browse 2002). Ebenso wird der Lipidanteil einiger bakterieller Glycolipide von iterativen PKS-I bereitgestellt (Campbell, Cohen und Meeks 1997). Diese Gruppe enthält u. a. die Synthase des Heterocysten-Glycolipids Stickstoff-fixierender Cyanobakterien. Das gemeinsame Merkmal dieser beiden Enzym-Typen ist eine besondere Domänen-Architektur der Form KS-AT-[ACP]₂₋₅-(KR), d. h. sie weisen zwischen zwei bis fünf ACP-Domänen auf sowie eine optimale KR-Domäne, die untypischerweise hinter dem ACP liegt. Dies trifft nur auf die beiden actinobakteriellen Proteine SgcE aus *Streptomyces globisporus* und NcsE aus *S. carzinostaticus* nicht zu. Beide sind an der Biosynthese von Enediyn-Antibiotika beteiligt, SgcE an der Synthese von C-1027 (Liu *et al.* 2002) und NcsE an der Synthese von Neocarzinostatin (Liu *et al.* 2005).

Als nächste Gruppe finden sich die modular aufgebauten PKS-I, wobei die PKS mit distinkten AT-Domänen (*trans*-AT) eine abgeschlossene phylogenetische Gruppe bilden, wie aus der Literatur bereits bekannt ist (Piel *et al.* 2004). Sie enthält viele Sequenzen aus *Bacillus subtilis*, dessen PKS-I ausschließlich diesem Typ angehören, aus dem Leinamycin-Biosynthese-Cluster von *Streptomyces atrolivaceus* (Tang, Cheng und Shen 2004) und aus dem Pederin-Cluster des bakteriellen Symbionten des Käfers *Paederus fuscipes* (Piel 2002). Daneben steht die Gruppe der PKS-I mit integrierter *cis*-AT. In dieser liegt auch ein Zweig mit der iterativen Orsellinsäure-Synthase AviM der Avilomycin-Biosynthese (Weitnauer *et al.* 2001) und dem sehr ähnlichen Enzym NcsB, das vermutlich den Naphtonsäure-Anteil von Neocarzinostatin bildet (Liu *et al.* 2005). Im selben Zweig ist die noch uncharakterisierte Pks4 aus *Streptomyces avermitilis* lokalisiert. Die KS-Domäne aus dem iterativ agierenden Modul AurA der Aureothin-Biosynthese (He und Hertweck 2003) liegt dagegen in der Gruppe modularer PKS-I aus Streptomyceten, ohne phylogenetische Auffälligkeiten zu zeigen.

An der Spitze des Baums liegt ein Cladus eukaryotischer Sequenzen, der sich aus zwei Unterguppen zusammensetzt. Zum einen liegen dort die iterativen PKS-I aus Pilzen, zum anderen die tierischen FAS-I von Mensch, Ratte, Fruchtfliege und des Nematoden *Caenorhabditis elegans* sowie des Protozoen *Dictyostelium discoideum*, eines Schleimpilzes. Die Pilzgruppe umfasst auch das Protein MchA aus dem Myxobakterium *Stigmatella aurantiaca*, das für die Bildung der aliphatischen Seitenkette der Myxochromide verantwortlich ist (Wenzel *et al.* 2005). Seine Position erklärt sich am besten über ein HGT-Ereignis. Es fällt auf, dass die PKS-I aus Pilzen aus phylogenetischer Sicht sehr weit von den FAS derselben Organismengruppe entfernt liegen und stattdessen eine direkte Nachbargruppe der eukaryotischen FAS-I bilden, entsprechend der Übereinstimmungen hinsichtlich der Domänenanordnung. Ebenso fällt auf, dass diese beiden eukaryotischen Synthese-Systeme mit den modularen PKS-I aus Bakterien über einen gemeinsamen Vorläufer eng verknüpft sind.

Aus dem phylogenetischen Baum in **Abb. 8** ist klar erkennbar, dass die unterschiedlichen Typen von FAS und PKS in evolutionärer Hinsicht eng verflochten sind. Um weitergehend zu überprüfen, ob die Beziehungen zwischen modularen PKS-I aus Bakterien und der eukaryotischen Gruppe, bestehend aus tierischen FAS-I und iterativen PKS-I aus Pilzen, konsistent über die Gesamt-Systeme nachvollzogen werden können, wurden mit verkleinerten Datensätzen die entsprechenden Beziehungen auf Basis der anderen Domärentypen ebenfalls phylogenetisch untersucht. Dies geschah auch vor dem Hintergrund, dass die evolutionäre Entwicklung der tierischen FAS nach wie vor nicht abschließend geklärt ist.

In diese Datensätze wurden Sequenzen von PKS-I-Modulen verschiedener bakterieller Hauptgruppen aufgenommen, die neben den Basisdomänen auch die optionalen Domänen, d.h. DH-, ER- und KR-Domänen, enthalten. Die mit Hilfe des Bayesschen Verfahrens erhaltenen Bäume sind für die KS-, AT- und KR-Domänen in **Abb. 9** zusammengestellt, für die DH- und ER-Domänen in **Abb. 10**. Für alle Domärentypen wurden dieselben phylogenetischen Beziehungen gefunden, wie sie sich bereits aus dem größeren Datensatz der

Abb. 8 ergaben. Insbesondere weisen die tierischen FAS-I in allen Fällen einen nächsten gemeinsamen Vorläufer mit den PKS-I aus Pilzen auf, und diese Gruppe ist mit den bakteriellen PKS-I wiederum über einen gemeinsamen Verläufer eng verbunden. Für FAS-I aus Bakterien und Pilzen konnten die DH- und ER-Domänen nicht in einen gemeinsamen Datensatz mit den anderen Systemen analysiert werden, da diese Domärentypen einer anderen Familie angehören (Jenni *et al.* 2006) und die entsprechenden Sequenzen daher zu weit entfernt sind. So enthält die ER-Domäne von Pilz-FAS-I im Gegensatz zu allen anderen FAS-Systemen Flavinmononucleotid (FMN) als Cofaktor (Fox und Lynen 1980).

Aus den bislang vorgestellten Ergebnissen dieses Abschnitts wird erkennbar, dass die Frage nach dem evolutionären Ursprung der FAS aus Tieren prinzipielle Konstellationen der Evolution eukaryotischer Organismen berührt. Im Zusammenhang der frühen Eukaryoten-Evolution spielt innerhalb der Domäne Eukarya neben den gut aufgelösten Reichen der Pflanzen, Pilze und Tiere auch das Reich der Protisten eine wichtige Rolle. Es existieren zwar zahlreiche Sequenzierungsprojekte zu diversen Protistengruppen, allerdings war der Datenbestand zum Zeitpunkt der hier vorgestellten Untersuchung (September 2006) relativ gering. Mit der Proteinsequenz der FAS-I aus *Rattus norvegicus* und den Sequenzen der bakteriellen Proteine der Fettsäuresynthese als Suchsequenzen wurden die zur Verfügung stehenden Datenbanken nach ähnlichen Sequenzen mittels mehrerer BlastP-Analysen durchsucht. Die Ergebnisse sind in **Tab. 4** zusammengestellt. In zehn Genomen von Protisten konnten passende Sequenzen identifiziert werden. Überwiegend handelt es sich dabei um FAS vom Typ II, wobei sich zwar eindeutig KS und KR auffinden ließen, jedoch keine AT, DH und ER. Nur in zwei Fällen, nämlich in den Genomen von *Cryptosporidium hominis* und *Dictyostelium discoideum*, kommen auch FAS vom Typ I vor. Diese beiden Organismen verfügen darüber hinaus auch über PKS-ähnliche Enzymsysteme, wie in der Literatur bereits beschrieben wurde (Zhu *et al.* 2002; Xu *et al.* 2004; Eichinger *et al.* 2005). Biochemische Charakterisierungen dieser Enzyme wurden jedoch noch nicht veröffentlicht.

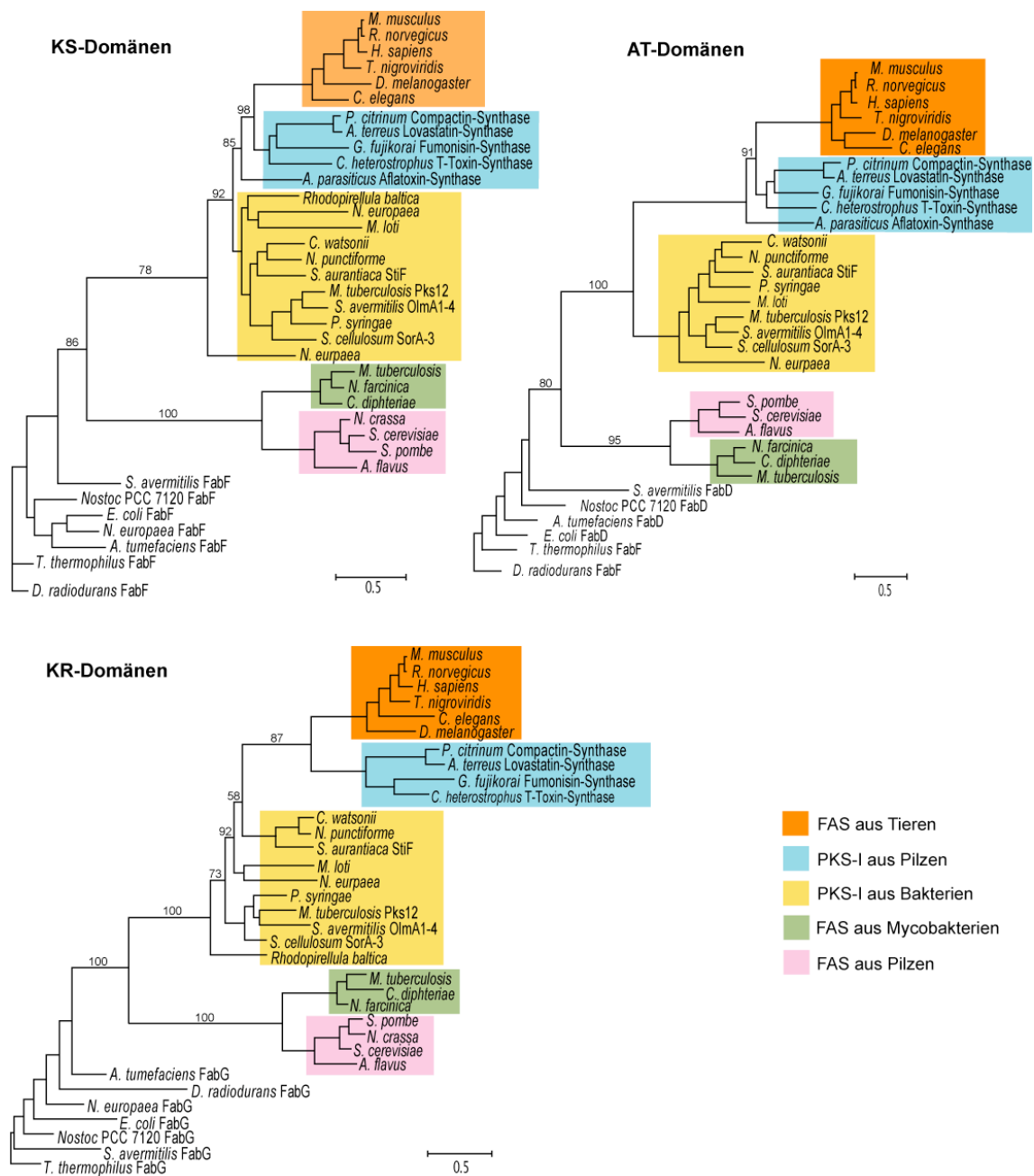


Abb. 9: Phylogenetische Bäume von KS-, AT- und KR-Domänen aus FAS und PKS im Vergleich.

Die Domänen stammen aus ausgewählten FAS und PKS der verschiedenen Organismengruppen. Die Enzymtypen und ihre Herkunft sind farblich unterlegt. Die Bäume wurden mit dem Bayes-Verfahren berechnet. An den Hauptverzweigungen der Bäume sind die posterioren *Clade credibility values* angegeben.

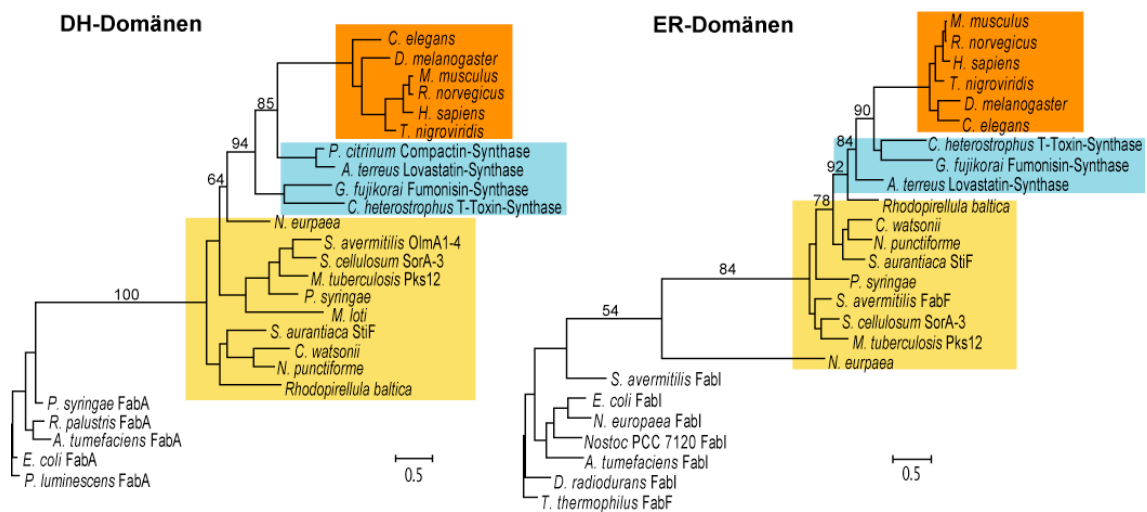


Abb. 10: Phylogenetische Bäume von DH- und ER-Domänen aus FAS und PKS zum Vergleich.

Die Domänen stammen aus ausgewählten FAS und PKS der verschiedenen Organismengruppen. Die Enzymtypen und ihre Herkunft sind farblich unterlegt. Die Farbgebung ist die gleiche wie in **Abb. 9**. Die Bäume wurden mit dem Bayes-Verfahren berechnet. An den Hauptverzweigungen der Bäume sind die posterioren *Clade credibility values* angegeben.

Die gefundenen Sequenzen wurden mit einem Teil des **Abb. 8** zugrunde liegenden Datensatzes kombiniert und einer phylogenetischen Analyse mit Hilfe des Bayes-Verfahrens unterzogen. Der resultierende phylogenetische Baum in **Abb. 11** zeigt, wie sich die Proteine aus Protisten in den bereits dargestellten Zusammenhang einordnen. Die KS aus *Plasmodium falciparum* und *Thalassiosira pseudonana* entstammen der plastidären Linie, was sich aus einer sekundären Endosymbiose einer bereits eukaryotischen Alge erklärt, die alle Organismen der Gruppe „Chromalveolaten“ betrifft (Cavalier-Smith 1998). Einige weitere Sequenzen fallen in das Cluster der mitochondrialen FAS-Systeme, darunter die Kinetoplastiden *Trypanosoma brucei* und *Leishmania major*, die keine anderen FAS zu besitzen scheinen. Dieselbe Situation trifft auf den Ciliaten *Tetrahymena thermophila* zu. Auch hier fügt sich die einzige auffindbare KS-Sequenz in den mitochondrialen Zweig des Baums ein. Ein aufgrund des sehr langen Zweiges und der damit verbundenen Schwierigkeiten im Hinblick auf phylogenetische Rekonstruktionen schwer zu beurteilendes kleines Cluster wird von den Sequenzen aus *Giardia lamblia* und *Entamoeba histolytica* gebildet, was insofern erstaunlich ist, als die entsprechenden Protistengruppen, Diplomonaden bzw. Amoebozoa, nach neueren Erkenntnissen als evolutionär relativ weit entfernt angesehen werden (Keeling *et al.* 2005). Schließlich fallen zwei große Cluster ins Auge, die auf FAS und PKS-ähnliche Proteine aus *Cryptosporidium hominis* und *Dictyostelium discoideum* zurückzuführen sind. Während die Sequenzen aus *C. hominis* in deutlicher Distanz zu den FAS-I und PKS-I aus Tieren bzw. Pilzen liegen, gruppieren alle Sequenzen aus *D. discoideum* zusammen mit der Sequenz von dessen FAS und somit in unmittelbarer Nähe zu den gut umschriebenen eukaryotischen Gruppen.

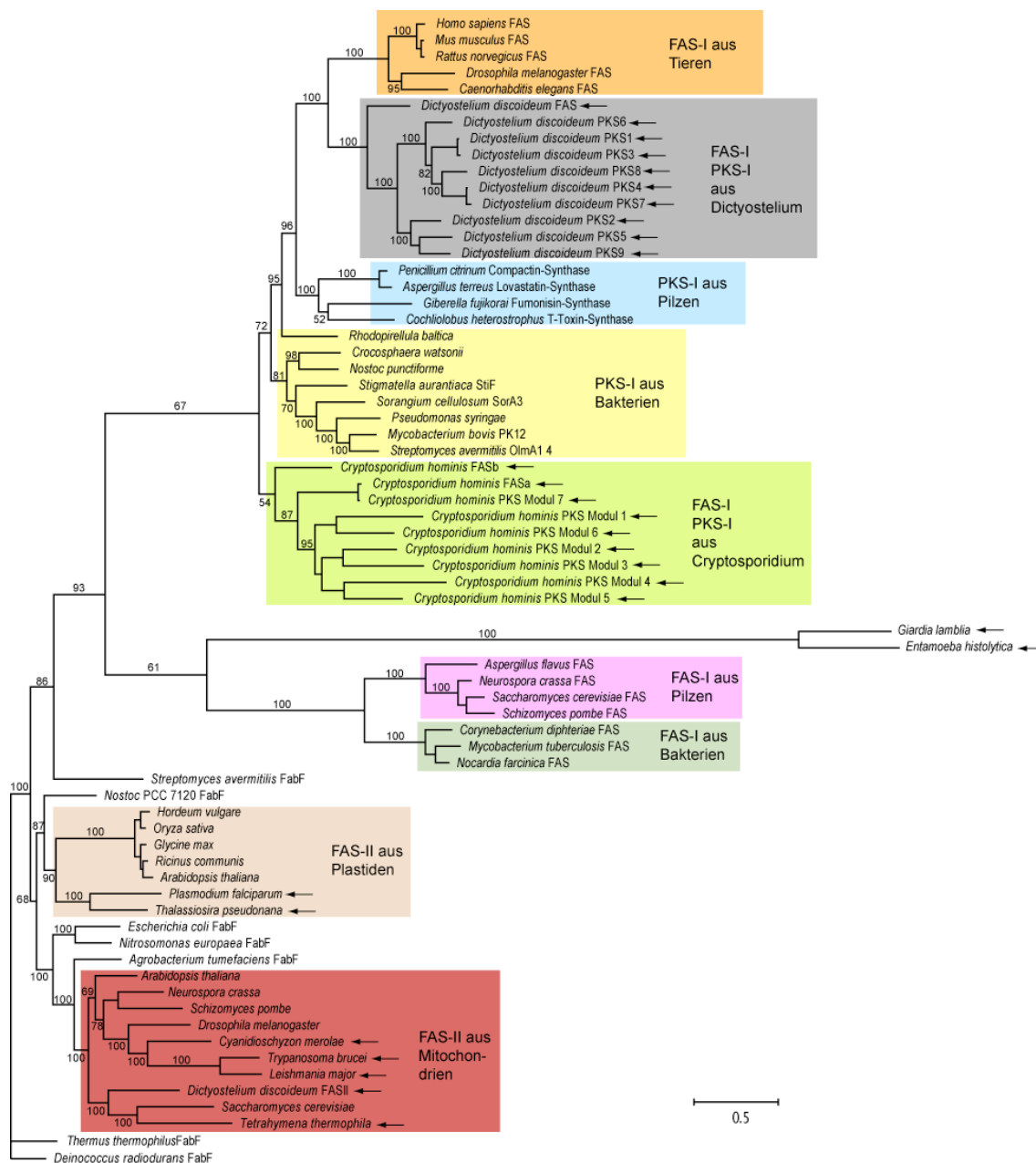


Abb. 11: Phylogentischer Baum der KS-Domänen bzw. -Proteine von FAS und PKS unterschiedlicher Organismenreiche unter Einschluss von Sequenzen aus Protisten.

Die Domänen stammen aus ausgewählten FAS und PKS der verschiedenen Organismengruppen. Die Enzymtypen und ihre Herkunft sind farblich unterlegt. Sequenzen aus Protisten sind durch einen Pfeil kenntlich gemacht. Der Baum wurde mit dem Bayes-Verfahren auf Basis der Aminosäure-Sequenzen berechnet. An den Hauptverzweigungen der Bäume sind die posterioren *Clade credibility values* angegeben. Die Längen der Zweige bezeichnen die durchschnittliche Zahl an Aminosäure-Austauschen pro Position.

Die evolutionären Zusammenhänge zwischen PKS und FAS werden in den Abschnitten 4.1.1 und 4.1.2 eingehend diskutiert.

Tab. 4: Protistengenome mit identifizierten Enzymen der Fettsäuresynthese

Organismus	Protistengruppe	FAS-Typ
<i>Plasmodium falciparum</i>	Apicomplexa	II
<i>Cryptosporidium hominis</i>	Apicomplexa	I
<i>Trypanosoma brucei</i>	Kinetoplastea	II
<i>Leishmania major</i>	Kinetoplastea	II
<i>Cyanidioschyzon merolae</i>	Cyanidiophyta	II
<i>Tetrahymena thermophila</i>	Ciliates	II
<i>Thalassiosira pseudonana</i>	Diatomea	II
<i>Giardia lamblia</i>	Diplomondia	II
<i>Entamoeba histolytica</i>	Lobosa	II
<i>Dictyostelium discoideum</i>	Amoebozoa	I und II

3.1.2 Phylogenetische Analyse von PKS-III aus verschiedenen Organismenreichen

Da PKS-III hinsichtlich ihres Aufbaus eine eigenständige Gruppe innerhalb der PKS bilden (siehe Abschnitt 1.2.4), wurden sie in einer eigenen phylogenetischen Analyse untersucht. Ursprünglich nahm man an, dass es sich um originär pflanzliche Biosynthese-Systeme handelt. Diese Sichtweise musste jedoch mit der Entdeckung dieser Enzyme in Bakterien (Moore und Hopke 2001) und Pilzen (Seshime *et al.* 2005) revidiert werden. Die Phylogenie wurde anhand eines Datensatzes aus Aminosäure-Sequenzen aller in öffentlichen Datenbanken abgelegten PKS-III aus Bakterien und Pilzen sowie einer Auswahl pflanzlicher PKS-III bestimmt. Die Rekonstruktion des phylogenetischen Baums erfolgte mit drei unterschiedlichen Methoden: dem Bayes-Verfahren, *Maximum Parsimony* und der Distanz-basierten *Neighbor-Joining*-Methode. Da alle Methoden zu fast identischen Baum-Topologien führten, ist nur das mit dem Bayes-Verfahren erhaltene Ergebnis gezeigt (**Abb. 12**).

Als Außengruppe wurden FabH-Sequenzen aus Archaeobakterien, Eubakterien und pflanzlichen Plastiden herangezogen, da sich PKS-III bekanntermaßen von der FabH-Komponente des bakteriellen Fettsäuresynthese-Wegs ableiten (Austin und Noel 2003). Die Sequenzen sind über drei bakterielle Gruppen, eine Pilz- und eine Pflanzengruppe verteilt, die alle durch hohe *Clade probability values* unterstützt werden. Die erste bakterielle Gruppe spaltet sich in zwei Untergruppen auf, von denen die erste RppA-ähnliche Sequenzen aus Streptomyceten sowie eine PKS-III aus dem Myxobakterium *Sorangium cellulosum* enthält. RppA aus *Streptomyces griseus* wurde als erste bakterielle PKS-III beschrieben und charakterisiert (Ueda *et al.* 1995; Funa *et al.* 1999). Sie katalysiert die Synthese von 1,3,6,8-Tetrahydroxynaphthalen aus fünf Malonyl-CoA-Einheiten. Die andere Untergruppe umfasst hauptsächlich Sequenzen von Bacilli und Mycobakterien. Darunter befindet sich auch die Pks11 aus *Mycobacterium tuberculosis*, die an der Synthese langkettiger α -Pyrone beteiligt ist (Saxena *et al.* 2003).

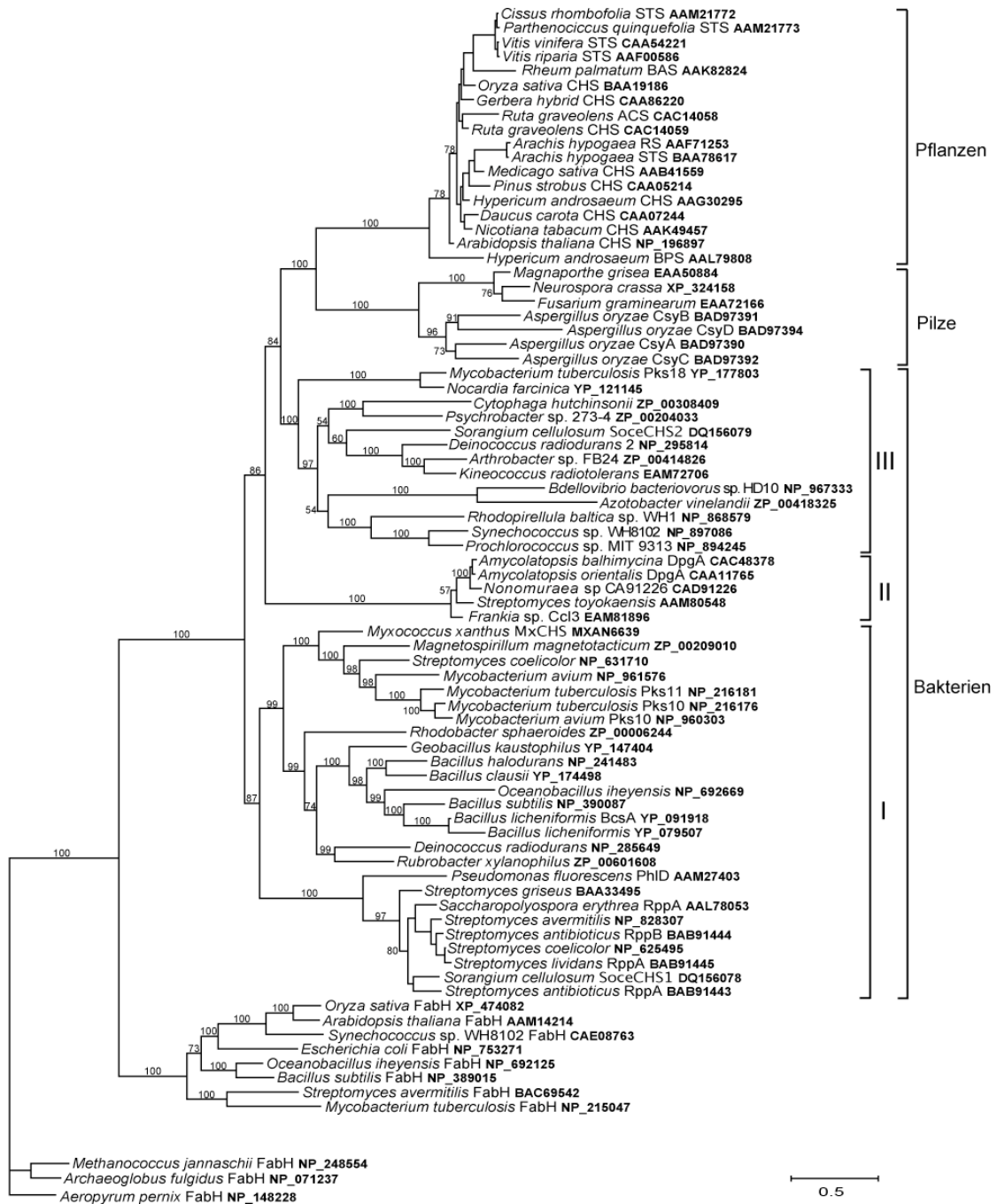


Abb. 12: Phylogenie der PKS-III aus Bakterien, Pilzen und Pflanzen.

Der Baum wurde mit dem Bayes-Verfahren berechnet. Die Zahlen über den Zweigen geben die *Clade probability values* an. Die verwendeten Sequenzen sind mit dem Spezies-Namen, der Bezeichnung des Proteins (falls annotiert) und der Zugriffsnummer ausgewiesen. Die Sequenzen aus Pflanzen sind der jeweiligen PKS-III-Untergruppe zugewiesen, wobei folgende Abkürzungen verwendet werden: CHS, Chalcon-Synthase; STS, Stilben-Synthase; BPS, Benzophenon-Synthase; BAS, Benzalaceton-Synthase; ACS, Acridon-Synthase; RS, Resveratrol-Synthase.

Der zweite bakterielle Cladus besteht ausschließlich aus actinobakteriellen PKS-III unter Einschluss des Enzyms DpgA aus *Amycolatopsis*-Spezies, das an der Synthese von Balhimycin, eines Antibiotikums der Vancomycin-Gruppe, mitwirkt (Pfeifer *et al.* 2001). Der dritte bakterielle Cladus ist im Vergleich zu den beiden vorgenannten recht heterogen aufgebaut. Unter anderem enthält er Proteinsequenzen aus Actinobakterien, Sphingobakterien, Cyanobakterien sowie γ - und δ -Proteobakterien. Keines dieser Enzyme wurde bislang biochemisch charakterisiert. Im oberen Bereich des Baums finden sich alle eukaryotischen Sequenzen des Datensatzes, wobei diese in zwei gut aufgetrennte Untergruppen aufspalten. Der erste Zweig umfasst die Proteine aus Pilzen, der zweite Zweig sämtliche pflanzliche PKS-III, die eine sehr homogene Gruppe darstellen. Sie enthält alle verschiedenen Vertreter, welche die Untergruppen pflanzlicher PKS-III repräsentieren. Diese Baum-Topologie wurde auch mit der MP- und der NJ-Methode erhalten mit der einzigen Ausnahme, dass im Ergebnis der MP-Analyse die Positionen der beiden bakteriellen Gruppen I und II vertauscht waren.

Der phylogenetische Baum steht in sehr guter Übereinstimmung mit früheren Studien, die zum einen eine klare Trennung zwischen bakteriellen und pflanzlichen PKS-III berichteten (Moore und Hopke 2001), zum anderen zeigen konnten, dass die Sequenzen aus Pilzen eine eigene phylogenetische Gruppe darstellen (Seshime *et al.* 2005). Im Verlauf neuer Genom-Projekte wurden in jüngster Zeit immer wieder neue PKS-III-Sequenzen gefunden, insbesondere in bakteriellen Genomen und dort auch in Bakteriengruppen, in denen bislang keine PKS-III beschrieben waren. So wurden in den Genomen der beiden Myxobakterien-Spezies *Myxococcus xanthus* und *Sorangium cellulosum* Soce56 Vertreter dieser Enzymgruppe identifiziert (mitgeteilt und Sequenzen zur Verfügung gestellt von Prof. R. Müller, Saarbrücken).

3.2 Evolutionäre Zusammenhänge zwischen bakteriellen modularen PKS-I

Modular aufgebaute PKS-I sind in Bakterien weit verbreitet. Diese Biosynthese-Systeme sind aus sehr großen Proteinkomplexen aufgebaut (siehe Abschnitt 1.2.2.1) und hauptsächlich in Actinobakterien, Myxobakterien, Pseudomonaden und Cyanobakterien beschrieben worden (Bode und Müller 2005). Systematische Studien zu ihrer Evolution unter Einbeziehung einer möglichst großen Zahl an Sequenzen wurden allerdings bislang noch nicht veröffentlicht. Um die Verteilung modularer PKS-I in den verschiedenen Bakteriengruppen und die Prinzipien, denen ihre Evolution unterliegt, untersuchen zu können, wurden zunächst alle veröffentlichten, vollständig sequenzierten Bakteriengenome nach entsprechenden Sequenzen durchsucht, diese Ergebnisse statistisch ausgewertet und im Folgenden auf der Basis dieser identifizierten PKS-I umfassende Phylogenien bezüglich der KS- und der AT-Domänen erstellt.

3.2.1 Verteilung modularer PKS-I innerhalb der Bakteriengruppen

Mit Hilfe der „Microbial Genome Platform“, die vom „National Center of Biotechnology Information“ (NCBI) zur Verfügung gestellt wird, wurde eine BlastP-Suche gegen 138 vollständige eubakterielle, 20

vollständige archaebakterielle und drei weitgehend fertig gestellte eubakterielle Genome (Stand: August 2005) durchgeführt, wobei die Sequenz von DEBS1 aus *Saccharopolyspora erythraea* als Suchsequenz diente. Nach manueller Überprüfung der Suchergebnisse ergab sich, dass kein archaebakterielles Genom PKS-I-Sequenzen enthielt, während 27 der insgesamt 141 eubakteriellen Genome positiv waren, was einem Anteil von etwa 20 % entspricht. Die Zahl der Genome mit PKS-I-Genen variierte beträchtlich zwischen den Bakteriengruppen. Die zur Verfügung stehenden Genome von Chlamydien und Spirochaeten enthielten keinerlei PKS-I-Sequenzen, der Anteil der positiven Genome in den Firmicuten, Proteobakterien, Cyanobakterien und Actinobakterien lag weit gestreut zwischen 5 % und 77 %. In **Tab. 5** sind die Ergebnisse für die vollständig sequenzierten Genome zusammenfassend dargestellt.

Tab. 5: Verteilung der in vollständig sequenzierten Bakteriengenomen codierten modularen PKS-I-Proteine

Bakteriengruppe	Zahl der Genome	Positive Genome	Zahl der PKS-I-Proteine	Zahl der Module	Module je Protein
Actinobakterien (gesamt)	13	10	98	132	1-5
Streptomyceten	2	2	25	55	1-5
Mycobakterien	5	5	70	74	1-2
Andere	6	3	3	3	1
Chlamydien	7	0	---	---	---
Cyanobakterien	11	5	34	35	1-2
Firmicuten	39	2	6	13	1-3
Proteobakterien (gesamt)	60	13	37	60	1-6
α -Proteobakterien	11	3	3	3	1
β -Proteobakterien	8	4	4	4	1
γ -Proteobakterien	33	5	11	12	1-2
δ -Proteobakterien	2	0	---	---	---
ϵ -Proteobakterien	5	0	---	---	---
Spirochaeten	4	0	---	---	---
Andere	8	1	2	2	1
Summe	141	30	177	242	---

Die überwiegende Mehrzahl der PKS-I-Proteine bestand aus einem einzelnen PKS-Modul, das zumindest die Basis-Konfiguration KS-AT-ACP zeigte. Multimodulare PKS-Proteine, d. h. solche, die mindestens zwei Module auf derselben Polypeptid-Kette tragen, wurden in großer Zahl in den Genomen von *Streptomyces avermitilis* und *Bacillus subtilis* (Stamm 168) gefunden. *B. subtilis* hob sich von allen anderen einbezogenen Bakterien-Spezies dadurch ab, dass er ausschließlich PKS-Module ohne AT-Domänen enthält. Eine vollständige Liste der Domänen-Architektur aller gefundenen PKS-I befindet sich im Anhang dieser Arbeit. Die meisten der für PKS-I codierenden Genome wiesen auch Sequenzen für NRPS auf. In Actinobakterien lagen PKS-I- und NRPS-Gene typischerweise in getrennten Gen-Clustern, während in Proteobakterien und Cyanobakterien häufig hybride PKS-I/NRPS-Systeme auftraten.

Von den 13 Genomen, die mindestens drei PKS-I-Gene enthalten, gehören sieben zur Gruppe der Actinobakterien, vier zu den Cyanobakterien, eines zu den *Bacillales* und eines zu den Pseudomonaden. Einschränkung muss berücksichtigt werden, dass die vorgestellten Zahlen insofern statistisch verzerrt sind, als bestimmte Bakteriengruppen unterrepräsentiert sind. Beispielsweise waren keine myxobakteriellen Genome in den Datenbanken vorhanden, es ist jedoch bekannt, dass diese Gattung viele von PKS-I synthetisierte Metabolite enthält (Bode und Müller 2005). Dennoch stimmen die erhaltenen Ergebnisse gut mit den Daten über die Zahl charakterisierter Verbindungen überein, die in den verschiedenen Gruppen beschrieben wurden. **Tab. 6** gibt einen Überblick über Verbindungen, die den modularen PKS-I aus Bakterien mit vollständig sequenzierten Genomen, zugeordnet werden können. Auch hier zeigen sich große Unterschiede zwischen den Bakteriengruppen. Im Fall von *Streptomyces avermitilis* können 16 der insgesamt 22 Proteine bekannten Polyketid-Verbindungen zugeordnet werden. Bei *Nostoc punctiforme* gilt dies hingegen nur für ein einziges PKS-I-Protein der 22 gefundenen Sequenzen. Es ist also beim derzeitigen Kenntnisstand unklar, wie viele der identifizierten Gene tatsächlich funktionell sind und wie viele der entsprechenden Enzyme konstitutiv oder nur unter bestimmten Umweltbedingungen synthetisiert werden.

Tab. 6: Namen der in der Analyse verwendeten Spezies und PKS-I-Proteine

Bakteriengruppe	Organismus	Zahl der Proteine	Proteinnamen	Bekannte Metabolite
Actinobakterien	<i>Corynebacterium diphtheriae</i>	1	---	keine
	<i>Corynebacterium efficiens</i> YS-314	1	---	keine
	<i>Corynebacterium glutamicum</i>	1	---	keine
	<i>Mycobacterium bovis</i> AF2122/97	18	PpsA-PpsE, Mas, Pks1-Pks9, Pks12, Pks13, Pks17	Phenolphthiocerol, Mycocerosinsäure
	<i>Mycobacterium leprae</i>	9	Mas, pks1	Mycocerosinsäure
	<i>Mycobacterium tuberculosis</i> CDC1551	17	PpsA-PpsE, Mas, Pks1-Pks9, Pks12, Pks13, Pks15, Pks17	Phenolphthiocerol, Mycocerosinsäure
	<i>Mycobacterium tuberculosis</i> H37Rv	17	PpsA-PpsE, Mas, Pks1-Pks9, Pks12, Pks13, Pks15, Pks17	Phenolphthiocerol, Mycocerosinsäure
	<i>Streptomyces avermitilis</i> MA-4680	22	AveA1-AveA4, OlmA1-OlmA7, PteA1-PteA5, Pks1-2, Pks1-3, Pks3-2, Pks4, Pks5, Nrps7	Avermectin, Oligomycin, Polyen-Makrolid
	<i>Streptomyces coelicolor</i> A3(2)	3	---	keine
Cyanobakterien	<i>Anabaena</i> sp. PCC 7120	4	---	keine
	<i>Nostoc punctiforme</i>	22	NosB	Nostopeptolide
	<i>Gloeobacter violaceus</i>	2	---	keine
	<i>Crocospaera watsonii</i>	3	---	keine
	<i>Anabaena variabilis</i>	3	---	keine

ERGEBNISSE

Firmicuten	<i>Clostridium acetobutylicum</i>	1	PksE	keine
α-Proteobakterien	<i>Mesorhizobium loti</i>	1	---	keine
	<i>Rhodopseudomonas palustris</i> CGA009	1	---	keine
	<i>Agrobacterium tumefaciens</i> C58	1	---	keine
β-Proteobakterien	<i>Bordetella bronchiseptica</i>	1	---	keine
	<i>Bordetella parapertussis</i>	1	---	keine
	<i>Ralstonia solanacearum</i>	2	---	keine
	<i>Nitrosomonas europaea</i>	1	---	keine
γ-Proteobakterien	<i>Escherichia coli</i> CFT073	3	---	keine
	<i>Photobacterium luminescens</i>	2	---	keine
	<i>Yersinia pestis</i> CO92	1	Irp1	Yersiniabactin
	<i>Yersinia pestis</i> KIM	1	Irp1	Yersiniabactin
	<i>Coxiella burnetii</i>	1	---	keine
	<i>Pseudomonas syringae</i> pv. tomato str. DC3000	3	Irp2	keine
δ-Proteobakterien	<i>Stigmatella aurantiaca</i>	14	StiA-StiJ, MxaB-MxaF	Stigmatellin, Myxalamid
	<i>Sorangium cellulosum</i> So ce90	5	EpoA-EpoE	Epothilon
	<i>Sorangium cellulosum</i> So ce26	2	SorA, SorB	Soraphen
Andere	<i>Rodopirellula baltica</i>	2	---	keine

3.2.2 Zusammenhang zwischen der Zahl an PKS-I-Genen und der Genomgröße

Wie im voran gegangenen Abschnitt dargestellt, variiert die Zahl an PKS-I-Genen zwischen Bakterien-Spezies erheblich. Ebenso gibt es große Unterschiede in den Genomgrößen der für die Untersuchung herangezogenen Spezies. Daher wurde überprüft, inwiefern eine statistisch signifikante Korrelation zwischen diesen beiden Größen besteht. **Abb. 13** zeigt die Zahl der Module von PKS-I, die von einem individuellen Bakteriengenom codiert werden, in Abhängigkeit der Genomgröße. Der Test auf eine nicht-parametrische Spearman-Korrelation ergab einen Korrelationskoeffizienten von $r=0,476$, wobei das 95 %-Konfidenzintervall zwischen 0,33 und 0,60 lag und der P-Wert kleiner als 0,0001 war, was eine statistisch signifikante Korrelation zwischen der Zahl an PKS-I-Modulen, die in einem Bakteriengenom codiert sind, und seiner Größe bedeutet. Generell weisen Bakteriengenome mit einer Größe unter 2 Mbp keine PKS-I-Gene auf, während fast alle Bakterien mit einer Genomgröße von mindestens 7 Mbp PKS-I-Gene besitzen. Die einzige Ausnahme unter den großen Genomen ist dasjenige von *Bradyrhizobium japonicum*, in dessen Genom mit einer Größe von 9,1 Mbp sich keine PKS-I-Sequenzen identifizieren ließen. Die

nachgewiesene Korrelation beschreibt den grundsätzlichen Trend, dass PKS-I-Gene in relativ großen Genomen erhalten, dupliziert und diversifiziert werden, wohingegen sie in Bakterien mit reduziertem Genom gar nicht oder in nur sehr geringer Anzahl vorkommen. Allerdings wird aus **Abb. 13** auch ersichtlich, dass es einige Bakterien-Spezies mit mittleren Genomgrößen um 4 Mbp gibt, die eine große Zahl an PKS-I-Modulen codieren. Besonders auffällig in dieser Hinsicht sind die im Datensatz enthaltenen Mycobakterien. Diese durchweg pathogenen Bakterien zeigen zwar teilweise reduzierte Stoffwechselwege als Folge der Ausnutzung des Wirtsstoffwechsels (Vissa und Brennan 2001), haben die für ihren ausgeprägten Lipid-Stoffwechsel notwendigen PKS-Enzyme jedoch trotz der allgemeinen Reduktionsbestrebungen erhalten.

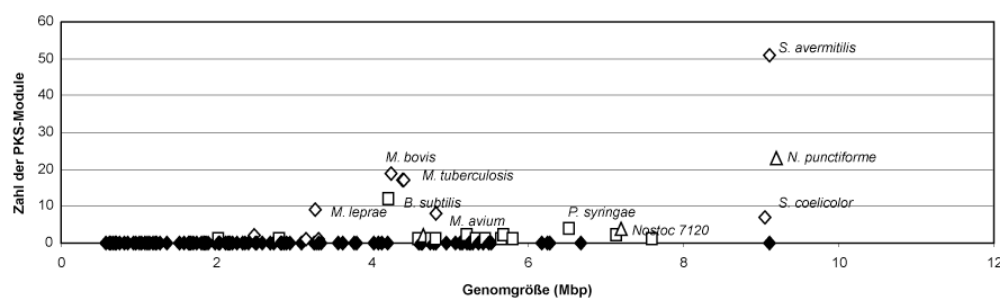


Abb. 13: Korrelation zwischen Genomgröße und Zahl der im Genom codierten Module von PKS-I.

Schwarze Rauten stehen für Genome ohne Gene für PKS-I. Die leeren Symbole bezeichnen Bakterienstämme, die über PKS-I-Gene verfügen. Actinobakterielle Stämme sind durch Rauten dargestellt, cyanobakterielle durch Dreiecke, alle übrigen durch Quadrate. Namen sind für alle Stämme angeführt, die mehr als drei PKS-I-Gene besitzen.

3.2.3 Phylogenetische Analyse modularer PKS-I

Die Beobachtung, dass modulare PKS-I in Bakterien weit verbreitet sind, jedoch in Bezug auf unterschiedliche Bakteriengruppen ein sehr heterogenes Verteilungsmuster zeigen, wirft die Frage auf, ob zwischen diesen Gruppen Unterschiede darin bestehen, wie sie PKS-I-Gene in ihr Genom integrieren und evolutionär weiter entwickeln. Um dieser Frage nachzugehen, wurden die identifizierten bakteriellen PKS-I phylogenetisch untersucht, wobei für die Analysen die Aminosäure-Sequenzen der ausgeschnittenen AT- und KS-Domänen verwendet wurden. Der auf den vollständig sequenzierten Genomen beruhende Datensatz wurde um die Enzyme von vier Biosynthesen aus Myxobakterien und einem Biosynthese-Weg aus einem Pseudomonaden ergänzt (vgl. **Tab. 6**), da innerhalb der Proteobakterien diese Gruppen angesichts der Vielzahl an bekannten Polyketid-Verbindungen deutlich unterrepräsentiert waren.

3.2.3.1 Phylogenie der AT-Domänen

Die phylogenetische Rekonstruktion basierte auf einem Aminosäure-Sequenz-Alignment mit 192 Sequenzen einschließlich von fünf FabD-Sequenzen bakteriellen Ursprungs, die als Außengruppe dienten.

Zur Rekonstruktion wurden mehrere Verfahren herangezogen: das Bayes-Verfahren, MP- und NJ-Methoden sowie mit kleineren Teildatensätzen eine ML-Analyse. Alle Verfahren ergaben sehr ähnliche Baum-Topologien, insbesondere im Hinblick auf alle Hauptverzweigungen. Unterschiede waren nur in der Anordnung sehr ähnlicher Sequenzen innerhalb von phylogenetischen Gruppen zu verzeichnen, weshalb nur der mittels des Bayes-Verfahrens errechnete Baum gezeigt ist (**Abb. 14**). Er weist zwei Hauptgruppen auf. Der eine Cladus umfasst die Untergruppen A1 bis A4, enthält also Malonyl-CoA-benutzende AT-Domänen (aufgrund biochemischer Charakterisierung bekannt oder anhand des aktiven Zentrums vorhergesagt), außerdem einige Domänen, deren vermutliche Substratspezifität nicht ermittelt werden konnte. Der zweite Cladus vereint solche AT-Domänen, die Methylmalonyl-CoA oder seltene Substrate verwenden (Untergruppen A6 bis A8) und eine Gruppe von Domänen, deren Substrat gesichert oder gemäß der Vorhersage Malonyl-CoA ist (Untergruppe A5). Eine völlig mit der hier gezeigten übereinstimmende Topologie ließ sich auch mit einem Sequenz-Alignment erhalten, aus dem die Aminosäure-Reste, die das aktive Zentrum bilden, entfernt worden waren (Daten nicht gezeigt). Dabei handelt es sich, bezogen auf das Protein FabD aus *Escherichia coli* um die Reste Q11, Q63, G90, H91, L93, G94, R117, S200, H201, N231, Q250 und V255 (Serre *et al.* 1995). Diese Reste können zur Vorhersage der Substratspezifität benutzt werden (Yadav, Gokhale und Mohanty 2003). Dass die AT-Domänen entsprechend ihrer Substratspezifität im phylogenetischen Baum aufspalten, selbst wenn diese Reste nicht in die Berechnung mit eingehen, ist von evolutionärer Bedeutung. Diese Aufspaltung spiegelt nämlich nicht nur die funktionelle Spezialisierung der enzymatischen Aktivitäten wider, sondern auch ihre evolutionäre Beziehung zueinander. Geht man davon aus, dass die AT-Domänen bakterieller PKS-I ursprünglich entsprechend ihrer Herkunft von FAS-Komponenten nur Malonyl-CoA als Substrat verwendeten, so ergibt sich die AT-Phylogenie als Ergebnis einer späteren Spezialisierung durch Änderung der Substratspezifität. Dabei erweisen sich die Domänen dieses Cladus als monophyletisch, d. h. sie haben sich nicht mehrfach unabhängig voneinander aus Vorläufer-Sequenzen entwickelt. Diese klare Trennung der beiden Hauptgruppen im AT-Baum zeigt sich auch anhand von Sequenzvergleichen. So weisen etwa die actinobakteriellen Sequenzen innerhalb eines Cladus eine Ähnlichkeit von durchschnittlich ca. 70 % auf, während dieser Wert beim Vergleich zwischen den Gruppen 50 % nicht überschreitet.

Abgesehen von den unterschiedlichen Substratspezifitäten, ließen sich keine Kriterien einer funktionellen Divergenz mit entsprechender Gruppenbildung erkennen. Die phylogenetischen Gruppen im AT-Domänen-Baum sind unabhängig von der Domänen-Zusammensetzung der Module, also z. B. von dem Vorhandensein einer KR-Domäne. Ebenso wenig hat die Existenz eines NRPS-Moduls auf der Donor- oder Akzeptorseite einen Einfluss auf die Position der Domänen im Baum.

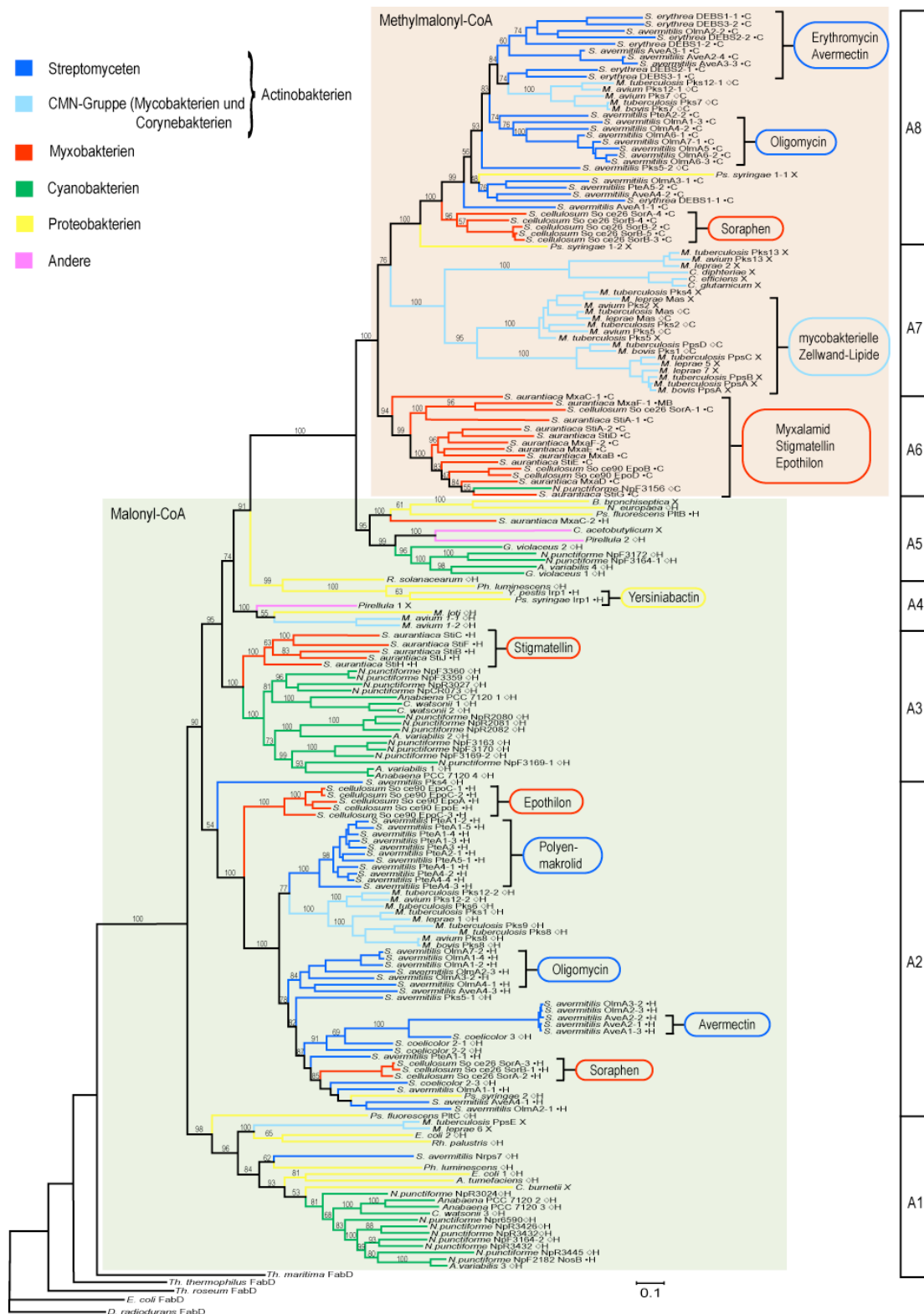


Abb. 14: Phylogenie der AT-Domänen bakterieller PKS-I mit modularem Aufbau.

Der Baum wurde mittels des Bayes-Verfahrens rekonstruiert. Oberhalb der Zweige sind die *Clade probability values* angegeben, die Längen der Zweige geben die Zahl der errechneten Aminosäure-

Fortsetzung der Legende zu Abb. 14

Austausche je Position an. Die Zweige sind entsprechend der Zugehörigkeit der Sequenz zu einer Bakteriengruppe farbig markiert, wie im Farb-Code links oben angegeben. AT-Domänen, die gemäß der Vorhersage der Substratspezifität oder gemäß ihrer biochemischen Charakterisierung Malonyl-CoA als Substrat verwenden, sind grün hervorgehoben, solche mit Methylmalonyl-CoA oder seltenen Substraten in orange. Die Enden der Zweige bezeichnen den Organismus, den Proteinnamen (falls in der Datenbank annotiert), die Modulnummer bei mehrmoduligen Proteinen und die Substratspezifität (H, Malonyl-CoA; C, Methylmalonyl-CoA; MB, Methylbutyryl-CoA; X, unklar). Biochemisch charakterisierte AT-Domänen sind durch schwarze Punkte gekennzeichnet. Domänen, deren Substratspezifität vorhergesagt wurde, sind mit Rauten markiert. Die farbig umrandeten Kästen mit Verbindungsnamen kennzeichnen phylogenetische Untergruppen von Domänen, die ausschließlich oder hauptsächlich an der Biosynthese dieser Verbindung beteiligt sind. Die Zahlen am Rand geben Gruppennummern an, die im Text verwendet werden.

3.2.3.2 Phylogenie der KS-Domänen

Wie für die AT-Domänen beschrieben, wurde die phylogenetische Rekonstruktion mit unterschiedlichen Methoden durchgeführt, um die Stabilität der erhaltenen Ergebnisse zu gewährleisten. Auch im Fall der KS-Domänen ergaben sich sehr ähnliche Baum-Strukturen, deren Unterschiede auf die letzten Verzweigungen innerhalb größerer Gruppen beschränkt waren. Gezeigt ist deshalb nur der Baum, der mit Hilfe des Bayes-Verfahrens gewonnen wurde (**Abb. 15**). Die phylogenetischen Verhältnisse zwischen KS-Domänen werden durch zwei Faktoren überlagert, die mit der funktionellen Umgebung der Domänen zusammenhängen. Der erste Faktor zeigt sich an der Gruppe K1, die Sequenzen aus allen im Datensatz vorkommenden Bakteriengruppen enthält. Alle KS-Domänen dieser Gruppe haben gemein, dass sie Bestandteil von PKS-I-Modulen sind, die innerhalb eines gemischten PKS/NRPS-Systems hinter einem NRPS-Modul liegen. Dies entspricht den Vorhersagen einer früheren Studie, in der dargelegt wurde, dass die biochemische Kommunikation zwischen den Modulen eines solchen Hybrid-Systems spezialisierte KS-Domänen erfordert, die in der Lage sind, die in reinen PKS-Systemen nicht auftretenden Peptidyl-Substrate von der Akzeptorseite des vorausliegenden Moduls als Substrat zu benutzen (Shen *et al.* 2001). Dass KS-Domänen aus gemischten PKS/NRPS-Multienzymen phylogenetisch in eigenen Gruppen zusammenfallen, wurde bereits in einer anderen Arbeit anhand eines deutlich kleineren Datensatzes festgestellt (Moffitt und Neilan 2003). Dieser Domänen-Typ scheint Sequenz-Besonderheiten aufzuweisen, die sich nicht allein auf das aktive Zentrum erstrecken, sondern die gesamte Domäne betreffen, und kann auf der hier verwendeten Datenbasis als monophyletisch klassifiziert werden.

Der zweite Faktor betrifft den im Vergleich zu Elongationsmodulen unterschiedlichen Aufbau der Beladungs-Module (siehe Abschnitt 1.2.2.1.2). Die Gruppe K2 enthält KS-Domänen aus solchen Beladungs-Modulen, die im Gegensatz zum ursprünglich beschriebenen Standard-Typ ebenfalls eine KS-Domäne besitzen. In diesen ist der essentielle Cystein-Rest im aktiven Zentrum häufig durch einen Glutamin-Rest ersetzt (Kao *et al.* 1996), weshalb sie auch als KSQ-Domänen bezeichnet werden. In der Gruppe K2 in **Abb. 15** befinden sich drei dieser besonderen KS-Domänen, die alle aus dem Streptomyceten *S. avermitilis* stammen. Sie liegen zusammen mit Sequenzen aus den Myxobakterien *Stigmatella aurantiaca* und *Sorangium cellulosum*, die zu so genannten Initiations-Doppelmodulen gehören. Diese Doppelmodule zeigen den Aufbau ACP-KS-AT-AT-DH-KR-ACP, wobei man davon ausgeht, dass die KS-Domäne ihre katalytische Aktivität im zweiten Modul entfaltet (Gaitatzis *et al.* 2001).

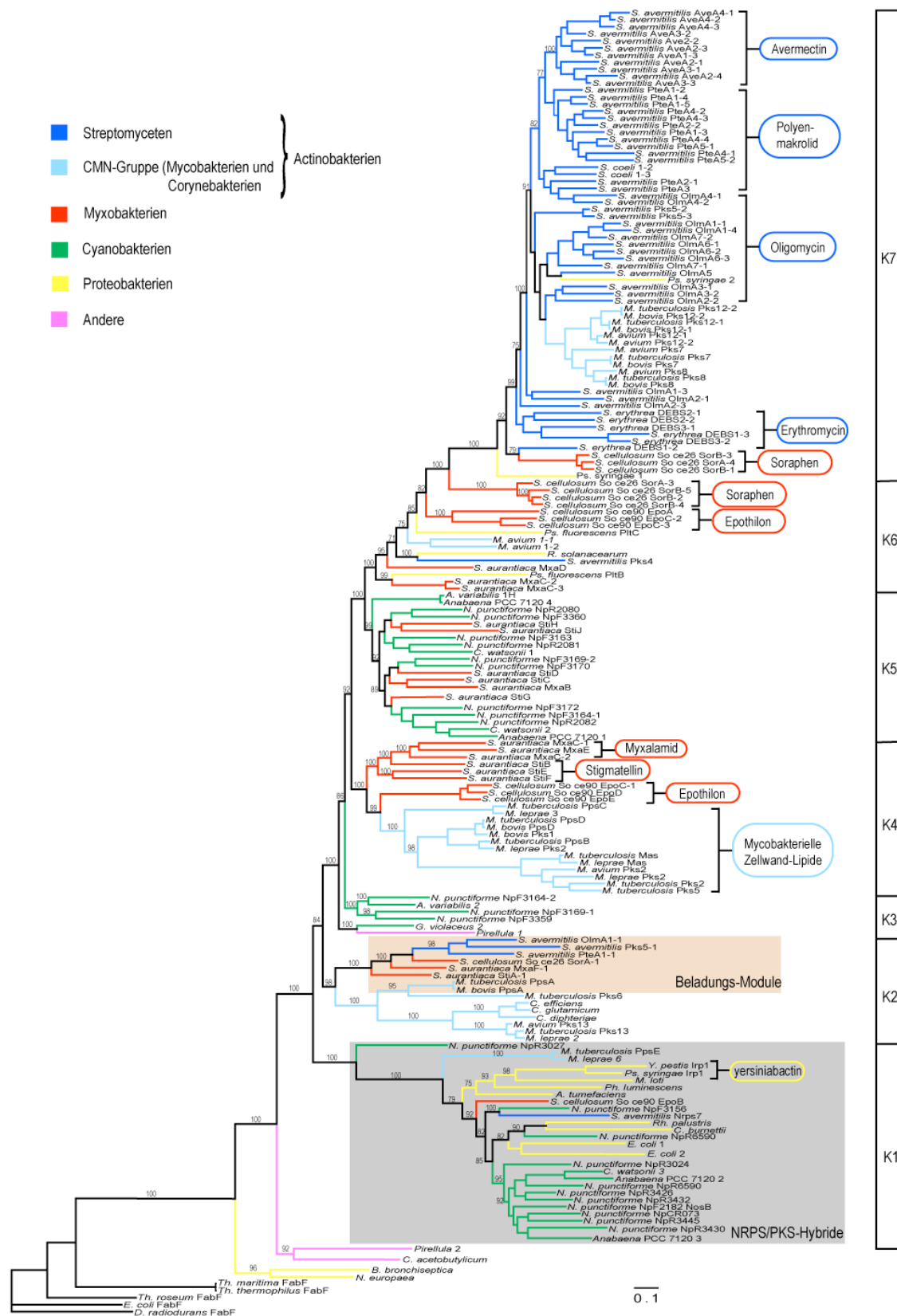


Abb. 15: Phylogenie der KS-Domänen bakterieller PKS-I mit modularem Aufbau.

Legende zu Abb. 15:

Der Baum wurde mit dem Bayes-Verfahren berechnet. Die Zahlen über den Zweigen geben die *Clade probability values* an. Die Längen der Zweige geben die Zahl der errechneten Aminosäure-Austausche je Position an. Die Zweige sind entsprechend der Zugehörigkeit der Sequenz zu einer Bakteriengruppe farbig markiert, wie im Farb-Code links oben angegeben. KS-Domänen, die Teil von hybriden NRPS/PKS-System sind, sind grau unterlegt. KS-Domänen, die zu Beladungs-Modulen gehören, sind braun unterlegt. Die farbig umrandeten Kästen mit Verbindungsnamen kennzeichnen phylogenetische Untergruppen von Domänen, die ausschließlich oder hauptsächlich an der Biosynthese dieser Verbindung beteiligt sind. Die Zahlen am Rand geben die im Text benutzten Gruppennummern an.

Wie sich aus dem Sequenz-Alignment ergibt, zeigen diese Domänen eine hohe Sequenz-Ähnlichkeit mit denen des KSQ-Typs, besitzen jedoch den normalen Cystein-Rest im aktiven Zentrum (Daten nicht gezeigt). Dasselbe gilt für das mycobakterielle Protein PpsA im Nachbarzweig des Baums. PpsA ist das erste Modul des Phthiocerol/Phenolphthiocerol-Biosynthese-Clusters, das an der Synthese mycobakterieller Zellwand-Lipide beteiligt ist (Azad *et al.* 1997).

Ansonsten zeigt der KS-Baum sehr ähnliche Muster phylogenetischer Beziehungen wie der AT-Baum in **Abb. 14**. Insbesondere ist zu erkennen, dass Domänen, die zum selben Biosynthese-Cluster gehören, in der ganz überwiegenden Zahl der Fälle auch in enger Nachbarschaft in gemeinsamen Gruppen liegen.

3.2.3.3 Bestimmung von Genduplikationen, Genverlusten und horizontalem Gentransfer aus den phylogenetischen Bäumen

Die Rekonstruktion der Phylogenien von AT- und KS-Domänen ermöglichte eine Analyse des Einflusses von Faktoren, die bekanntermaßen eine Rolle bei der evolutionären Umbildung von Genomen spielen. Zu diesen Faktoren gehören die Duplikation von Genen, der Verlust von Genen und der horizontale Gentransfer (HGT), d. h. die Aufnahme und Integration von Genmaterial anderer Organismen (Mira, Ochman und Moran 2001). Viele bakterielle Genome besitzen einen hohen Anteil an Genen, die Homologie zu anderen Genen innerhalb desselben Genoms zeigen (Tekaia und Dujon 1999). Für die Evolution modularer Enzym-Systeme, wie sie die hier betrachteten modularen PKS-I darstellen, sind Genduplikationen eine notwendige Voraussetzung. Die Erfassung von Genduplikationen, die zur Entstehung von Paralogen führen, basiert im Allgemeinen auf der Identifizierung homologer Gen-Sequenzen innerhalb eines Genoms, die sich in phylogenetischen Bäumen an entsprechenden Aufspaltungsmustern zeigen. Die Abschätzung der Zahl an Genduplikationen und Genverlusten anhand der phylogenetischen Bäume basierte auf folgenden Überlegungen. Sequenz-Cluster, die zur selben bakteriellen Gruppe und zumindest teilweise zum selben Organismus gehören, könnten bereits im gemeinsamen Vorläufer der im Cluster vertretenen Organismen vorhanden gewesen sein. Die Alternative besteht darin, dass sie erst nach Spezifikationsprozessen entstanden sind. Trifft die erste Annahme zu, so müssen Genverluste berücksichtigt werden, und die Zahl an Genduplikationen ergibt sich aus dem Organismus mit der größten Zahl an Genkopien. Der so berechnete Wert entspricht der zur Erklärung der Phylogenie minimal notwendigen Zahl an Genduplikationen. Bei der alternativen Annahme müssen hingegen keine Genverluste berücksichtigt werden, so dass sich die Maximalzahl an Genduplikationen direkt als Summe der Duplikationen in allen Organismen dieser phylogenetischen Gruppe ergibt.

Im Gegensatz zu Duplikations- und Verlustereignissen ist die Detektion von HGT deutlich schwieriger und mit größerer Unsicherheit behaftet (siehe hierzu unter anderem die ausführlichen Darstellungen bei Koonin, Makarova und Aravind 2001; Rangan, Joshi und Smith 2001; Philippe und Douady 2003). Die drei Hauptkriterien für HGT sind: Inkongruenzen in phylogenetischen Bäumen im Vergleich zu Spezies-Phylogenien, die anomale Verteilung von Genfamilien innerhalb von taxonomischen Gattungen und eine atypische Zusammensetzung der Sequenzen im Hinblick auf die Basenanteile und den Codon-Gebrauch (Ragan 2001), wobei eine möglichst gute Absicherung der Ergebnisse mittels Kombination dieser Kriterien erreicht werden kann. Um phylogenetische Inkongruenzen in den AT- und KS-Bäumen identifizieren zu können, wurde eine Spezies-Phylogenie derjenigen bakteriellen Arten erstellt, deren Sequenzen in die Domänen-Analyse eingingen. Die Spezies-Phylogenie wurde auf der Basis der DNA-Sequenzen der kleinen ribosomalen Untereinheit (16S-DNA) mit Hilfe des NJ-Verfahrens berechnet und ist in **Abb. 16** dargestellt.

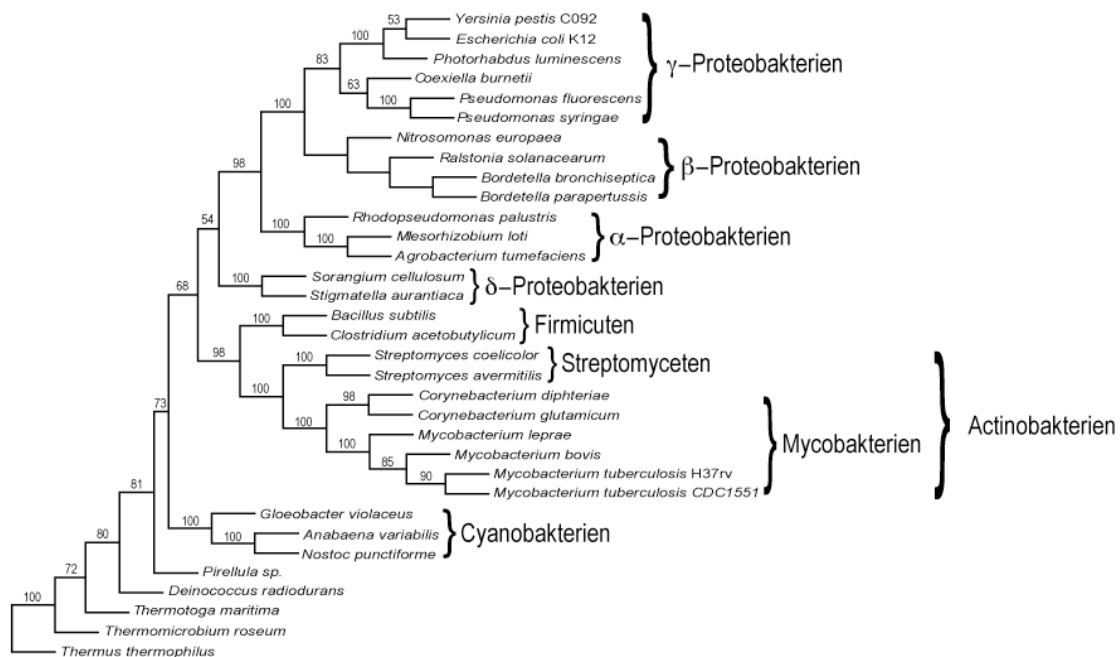


Abb. 16: Spezies-Phylogenie, basierend auf den Sequenzen der 16S-DNA.

Der phylogenetische Baum wurde mit Hilfe des NJ-Verfahrens erstellt. Erfasst wurden Bakteriengruppen, die in den Datensätzen für die phylogenetische Analyse der AT- und KS-Domänen modularer PKS-I vertreten waren. Die Zahlen über den Zweigen geben die *Bootstrap*-Werte an, die anhand von 1000 Pseudo-Sequenzreplikaten berechnet wurden.

Aus dem Baum ist ersichtlich, dass die Stämme innerhalb der bakteriellen Hauptgruppen (Cyanobakterien, Actinobakterien und Proteobakterien) monophyletischen Ursprungs sind. Dasselbe Ergebnis wurde auch mit den hoch konservierten Aminosäure-Sequenzen des Elongationsfaktors Tu und der weniger stark konservierten Sequenz der FAS-Komponente FabD erhalten (Daten nicht gezeigt). Diese Baum-Topologie stimmt überein mit phylogenetischen Rekonstruktionen, die anhand von anderen 16S-DNA-Datensätzen

(Woese 1987), Proteinen des Translationsapparates (Brochier *et al.* 2002) und diversen Kombinationen von Protein-Datensätzen (Daubin, Gouy und Perriere 2001; Battistuzzi, Feijao und Hedges 2004) erstellt wurden.

Der phylogenetische Baum der AT-Domänen (**Abb. 14**) zeigt, dass der Einfluss von Genduplikationen und potentiellen HGT-Ereignissen in den verschiedenen Bakteriengruppen recht unterschiedlich ausfällt. Nach den weiter oben dargestellten Prinzipien wurden die Ereignisse für die einzelnen Bakteriengruppen systematisch durchgezählt. Die Ergebnisse sind in **Tab. 7** zusammengefasst. Wertet man die Phylogenie der KS-Domänen aus, erhält man sehr ähnliche Ergebnisse (Daten nicht gezeigt).

Tab. 7: Geschätzte Anzahl von Genduplikationen, Genverlusten und HGT-Ereignissen, ermittelt aus dem phylogenetischen Baum der AT-Domänen

	Genduplikationen		Genverluste	HGT-Ereignisse
	Minimale Anzahl	Maximale Anzahl		
Actinobakterien	72	58	38	0
Streptomyceten	51	43	4	0
Mycobakterien	21	15	34	0
Andere	0	0	0	0
Cyanobakterien	21	18	19	1
Proteobakterien	24	19	15	9
(α , β und γ)	1	1	0	5
Myxobakterien	23	18	15	4

Diese Ergebnisse erlauben, die PKS-I-Gene enthaltenden Bakterien in drei Gruppen einzuteilen. Die erste Gruppe umfasst solche Bakterien, in denen die meisten oder alle PKS-I-Gene von gemeinsamen Vorläufern abstammen. Die zweite Gruppe wird von Bakterien gebildet, die ihre PKS-I-Gene über HGT-Ereignisse erhalten haben, ohne sie im weiteren Verlauf mittels Genduplikationen weiter zu entwickeln. Die dritte Gruppe schließlich stellt sich gewissermaßen als Mischung der beiden anderen Gruppen dar, d. h. in diesen Bakterien finden sich sowohl viele Duplikationen als auch die Erweiterung des Genbestands durch HGT. Actinobakterien und Cyanobakterien gehören zur ersten Gruppe, wobei der Umstand, dass sich fast alle Sequenzen auf gemeinsame Vorläufer innerhalb der Bakteriengruppen zurückführen lassen, vereinzelte HGT-Ereignisse innerhalb der jeweiligen Gruppe nicht ausschließt. Solche Austausche durch HGT innerhalb einer Gattung wurden jedoch durch die hier angewendete phylogenetische Methode nicht erfasst. In den Cyanobakterien findet sich eine Sequenz, für die es klare Anzeichen einer Entstehung durch HGT gibt, und zwar die Sequenz NpF3156 aus *Nostoc punctiforme*. Es handelt sich dabei um die einzige AT-Domäne aus Cyanobakterien im Datensatz, die gemäß der Vorhersage Methylmalonyl-CoA als Substrat benutzt. Alle anderen cyanobakteriellen AT-Domänen verwenden Malonyl-CoA. Die Sequenz NpF3156 fällt in eine phylogenetische Gruppe zusammen mit einer Reihe myxobakterieller Sequenzen,

was eine klare phylogenetische Inkongruenz darstellt. Allerdings konnten auf DNA-Ebene keine Auffälligkeiten hinsichtlich der Basenzusammensetzung oder der Codon-Benutzung festgestellt werden.

In die zweite Gruppe fallen Sequenzen von α -, β - und γ -Proteobakterien. Ihre Verteilung im phylogenetischen Baum zeigt deutlich, dass sie aus anderen Bakteriengruppen importiert wurden. Besonders auffällig ist die Position zweier AT-Domänen aus *Pseudomonas syringae* mit den Nummern 1 und 2 in **Abb. 14**. Sie sind in den Gruppen A2 bzw. A7 in großer Nähe zu Sequenzen aus Actinobakterien gruppiert. Beide gehören zum Biosynthese-Komplex der Coronofacin-Säure, der an der Synthese des Phytotoxins Coronatin beteiligt ist (Bender, Alarcon-Chaidez und Gross 1999). Die entsprechenden Nucleotid-Sequenzen besitzen einen GC-Gehalt von 68 %. Dieser Wert entspricht weitaus besser dem typischen hohen GC-Gehalt von Streptomyceten und weicht signifikant vom durchschnittlichen GC-Gehalt von Pseudomonaden im Bereich von etwa 58 % ab. In diesen Fällen fanden sich also deutliche Belege für einen HGT von einem Streptomyceten zu *P. syringae* oder einem Vorläufer. Die AT-Domäne des Proteins Irp1 aus *P. syringae* findet sich im phylogenetischen Baum in der Gruppe A4 in direkter Nachbarschaft zu Sequenzen eines Gen-Clusters aus *Yersinia pestis*, das an der Biosynthese des Eisen-Chelators Yersiniabactin beteiligt ist (Miller *et al.* 2002). Wie im Fall der beiden anderen erwähnten Sequenzen übersteigt der GC-Gehalt der entsprechenden Nucleotid-Sequenzen in *P. syringae* und *Y. pestis* mit 65 % bzw. 61 % deutlich die für diese Bakterien typischen Durchschnittswerte. Das Gen-Cluster für die Yersiniabactin-Biosynthese könnte also von einer anderen Bakteriengattung mit hohem GC-Gehalt, z. B. aus Actinobakterien, stammen und mittels HGT auf *P. syringae* und *Y. pestis* übertragen worden sein. Insgesamt finden sich in der Gruppe der Proteobakterien dreierlei Arten von Hinweisen auf HGT: anomale Verteilung der Gene, Inkongruenzen im Vergleich zur Spezies-Phylogenie und deutliche Abweichungen im GC-Gehalt der betreffenden Gene.

Myxobakterielle Sequenzen liegen zusammen mit solchen aus Actinobakterien oder Cyanobakterien. Im Verhältnis zwischen Myxobakterien und Actinobakterien ließen sich keine weiteren Schlussfolgerungen hinsichtlich potentieller HGT-Ereignisse und deren Richtung ableiten, da die Genome beider Bakteriengruppen durch hohen GC-Gehalt im Bereich von ca. 70 % gekennzeichnet sind. Im Gegensatz dazu weisen cyanobakterielle Genome im Allgemeinen einen deutlich niedrigeren GC-Gehalt von etwa 45 % auf. Die Untersuchung der Nucleotid-Sequenzen der für die in der phylogenetischen Gruppe A3 (**Abb. 14**) liegenden Proteine codierenden Gene erbrachte allerdings keine signifikanten Abweichungen von den jeweils typischen GC-Gehalten, so dass sich mögliche HGT-Ereignisse auf dieser Ebene nicht unterstützen ließen. Die Untersuchungen wurden im Hinblick auf die Myxobakterien dadurch eingeschränkt, dass noch keine vollständigen Genom-Sequenzen aus dieser Bakteriengruppe vorlagen.

3.2.3.4 Vergleich der phylogenetischen Bäume der AT- und KS-Domänen

Neben der Einzelauswertung der phylogenetischen Bäume der AT- und KS-Domänen, wie sie in den voran gegangenen Abschnitten dargelegt wurde, konnten die Bäume auch für einen Vergleich herangezogen werden, aus dem sich ableiten ließ, dass es zwischen den Modulen von PKS-I zu Austausch von Domänen kommt. Ausgangspunkt für diesen Vergleich war die Beobachtung, dass die evolutionäre Ausformung von PKS-I-Clustern nicht allein auf Genduplikationen beruhen kann, was sich bereits aus der Domänenstruktur einzelner Module ergibt. Beispielsweise umfasst das Protein AveA4, das an der

Biosynthese von Avermectin in *S. avermitilis* beteiligt ist, drei Module mit der folgenden Zusammensetzung: Modul 1: KS-AT-KR-ACP, Modul 2: KS-AT-ACP und Modul 3: KS-AT-DH-KR-ACP. Wenn sich diese Module durch Genduplikationen gebildet haben, so müssen zusätzlich der Verlust oder die Integration von DH- und KR-Domänen passiert sein, um die vorliegende Struktur zu erklären.

Bei einem Vergleich der phylogenetischen Bäume der AT- und KS-Domänen ließen sich einige Beispiele für solche Modul-Veränderungen auffinden. Das Protein OlmA6 aus der Biosynthese von Oligomycin stellt ein Beispiel für Module dar, die sich allein durch Genduplikationen entwickelt haben könnten. Alle drei Module, die auf diesem Protein integriert sind, zeigen den gleichen Aufbau, nämlich die Domänen-Abfolge KS-AT-KR-ACP. Außerdem setzen die AT-Domänen dieser Module alle das gleiche Substrat um. Die KS-Domänen liegen im Baum eng benachbart innerhalb der gleichen Untergruppe (**Abb. 15**, Gruppe K7), ebenso sind die AT-Domänen dieser Module phylogenetisch sehr eng verwandt (**Abb. 14**, Gruppe A8). Die unmittelbare Verwandtschaft dieser Module wird weiterhin unterstützt durch eine hohe Ähnlichkeit der entsprechenden Nucleotid-Sequenzen, die sich auch auf die Interdomänen-Bereiche erstreckt, die ansonsten eine relativ hohe Variabilität zeigen (Daten nicht gezeigt). Eine andere Situation findet sich bei Enzymen des Epothilon-Clusters von *Sorangium cellulosum*. Alle Malonyl-CoA-spezifischen AT-Domänen liegen in derselben Untergruppe (**Abb. 14**, Gruppe A2), während die dazu gehörigen KS-Domänen in zwei unterschiedlichen Gruppen lokalisiert sind (**Abb. 15**, Gruppen K4 und K6).

Solche Austauschprozesse konnten im Vergleich der Bäume auch bezüglich von AT-Domänen mit unterschiedlicher Substratspezifität festgestellt werden. Das Protein OlmA4 beispielsweise umfasst nur zwei Module, von denen eines Malonyl-CoA verwendet, das andere dagegen Methylmalonyl-CoA. Diese beiden AT-Domänen fallen phylogenetisch weit auseinander. Die eine liegt im unteren Teil des Baums (**Abb. 14**, Gruppe A2), die andere ihrer Substratspezifität gemäß im oberen Teil (**Abb. 14**, Gruppe A8). Dagegen sind die zugehörigen KS-Domänen eng verwandt, wie sich daran zeigt, dass sie in unmittelbarer Nachbarschaft zueinander liegen (**Abb. 15**, Gruppe K7).

Die genannten Beispiele wiesen darauf hin, dass Veränderungen der Architektur von PKS-I-Modulen nach erfolgter Genduplikation einen wichtigen Beitrag zur Evolution dieser Synthese-Systeme leisten. Deshalb wurde eine weiterführende Analyse solcher Veränderungen anhand eines dafür geeigneten Genoms durchgeführt.

3.3 Intragenomische Evolution modularer PKS-I

Zur Vertiefung der Analyse aus Abschnitt 3.2.3.4 wurde die intragenomische Evolution der PKS-I-Cluster in dem Streptomycet *Streptomyces avermitilis* untersucht. Dieses Genom wurde ausgewählt, da es aus folgenden Gründen besonders geeignet erschien: (1) Die vollständige Genom-Sequenz dieses Bakteriums wurde vor einiger Zeit veröffentlicht (Omura *et al.* 2001). (2) Dieses Genom ist unter allen bislang veröffentlichten Bakteriengenomen dasjenige, welches die meisten Gensequenzen für PKS-I enthält. (3) Die Mehrzahl der Module dieser PKS-I kann der Biosynthese von drei charakterisierten Verbindungen, Avermectin, Oligomycin und einem Polyenmakrolid, zugeordnet werden (Omura *et al.* 2001). Zunächst wurden die PKS-I-Proteine aus *S. avermitilis* zusammengestellt und anhand der KS-Domänen in einen phylogenetischen Gesamttzusammenhang mit anderen PKS-I-Clustern aus Streptomyceten gebracht.

Anschließend wurden die Einzelphylogenien der verschiedenen Domärentypen bestimmt, um die Module im Hinblick auf Veränderungen der Domänenzusammensetzung untersuchen zu können.

3.3.1 PKS-I-Cluster in *Streptomyces avermitilis*

Das Genom von *S. avermitilis* enthält codierende Sequenzen für acht PKS-I-Cluster (Omura *et al.* 2001). Die Cluster, die an der Biosynthese von Avermectin (*ave*), Oligomycin (*olm*) und dem Polyenmakrolid (*pte*) beteiligt sind, umfassen zwischen 80 und 100 kbp und stellen damit 86 % der insgesamt 51 PKS-Module. Die Strukturen der genannten Metabolite sind in **Abb. 17** dargestellt.

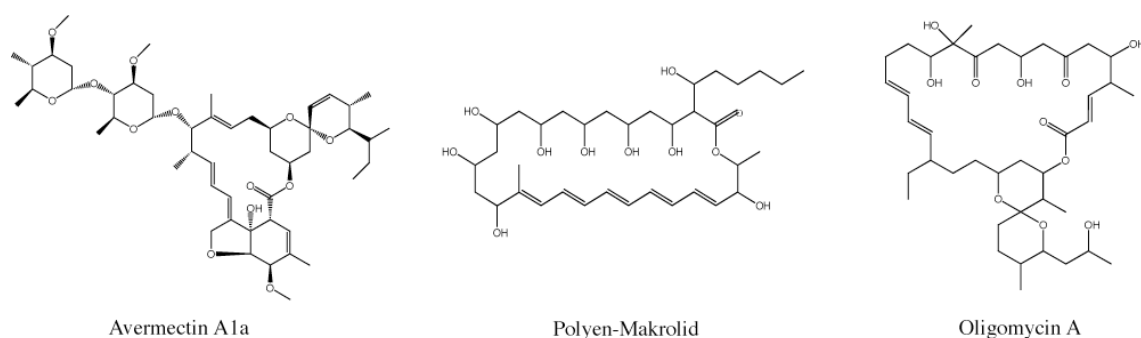


Abb. 17: Strukturen der von *S. avermitilis* mittels PKS-I produzierten Verbindungen

Die abgebildeten Strukturen sind Beispiele für die jeweiligen Substanzgruppen. Die gezeigte Struktur des Polyen-Makrolids ist ein Vorschlag der Arbeitsgruppe Omura und wurde der Website des Genom-Projekts entnommen.

Der Aufbau der einzelnen Proteine dieser drei großen PKS-I-Cluster hinsichtlich ihrer Modul- und Domänenstruktur ist in **Tab. 8** schematisch beschrieben.

Tab. 8: Modul- und Domänenstruktur der drei großen PKS-I-Cluster in *S. avermitilis*

Cluster	Protein	Modul	Domänenstruktur					
Avermectin	AveA1	AveA1-1	---	AT(MB)	---	---	---	ACP
		AveA1-2	KS	AT(CH3)	---	---	KR	ACP
		AveA1-3	KS	AT(H)	DH	---	KR	ACP
	AveA2	AveA2-1	KS	AT(H)	---	---	---	ACP
		AveA2-2	KS	AT(H)	---	---	KR	ACP
		AveA2-3	KS	AT(H)	---	---	KR	ACP
		AveA2-4	KS	AT(CH3)	DH	---	KR	ACP
	Ave3	AveA3-1	KS	AT(CH3)	DH	---	KR	ACP

ERGEBNISSE

		AveA3-2	KS	AT(H)	DH	---	KR	ACP
		AveA3-3	KS	AT(CH3)	DH	---	KR	ACP
	AveA4	AveA4-1	KS	AT(H)	---	---	KR	ACP
		AveA4-2	KS	AT(CH3)	---	---	---	ACP
		AveA4-3	KS	AT(H)	DH	---	KR	ACP
Oligomycin	OlmA1	OlmA1-1	KS*	AT(H)	---	---	---	ACP
		OlmA1-2	KS	AT(H)	---	---	KR	ACP
		OlmA1-3	KS	AT(CH3)	---	---	KR	ACP
		OlmA1-4	KS	AT(H)	DH	ER	KR	ACP
	OlmA2	OlmA2-1	KS	AT(H)	---	---	KR	ACP
		OlmA2-2	KS	AT(CH3)	---	---	KR	ACP
		OlmA2-3	KS	AT(H)	DH	---	KR	ACP
	OlmA3	OlmA3-1	KS	AT(CH3)	DH	ER	KR	ACP
		OlmA3-2	KS	AT(H)	DH	---	KR	ACP
	OlmA4	OlmA4-1	KS	AT(H)	DH	---	KR	ACP
		OlmA4-2	KS	AT(CH3)	DH	ER	KR	ACP
	OlmA5	OlmA5	KS	AT(CH3)	DH	---	KR	ACP
	OlmA6	OlmA6-1	KS	AT(CH3)	---	---	KR	ACP
		OlmA6-2	KS	AT(CH3)	---	---	KR	ACP
		OlmA6-3	KS	AT(CH3)	---	---	KR	ACP
	OlmA7	OlmA7-1	KS	AT(CH3)	---	---	KR	ACP
		OlmA7-2	KS	AT(H)	DH	---	KR	ACP
Polyenmakrolid	PteA1	PteA1-1	KS*	AT(H)	---	---	---	ACP
		PteA1-2	KS	AT(H)	---	---	KR	ACP
		PteA1-3	KS	AT(H)	DH	---	KR	ACP
		PteA1-4	KS	AT(H)	DH	---	KR	ACP
		PteA1-5	KS	AT(H)	DH	---	KR	ACP
	PteA2	PteA2-1	KS	AT(H)	DH	---	KR	ACP
		PteA2-2	KS	AT(CH3)	DH	---	KR	ACP
	PteA3	PteA3	KS	AT(H)	DH	---	KR	ACP
	PteA4	PteA4-1	KS	AT(H)	---	---	KR	ACP
		PteA4-2	KS	AT(H)	---	---	KR	ACP
		PteA4-3	KS	AT(H)	---	---	KR	ACP

	PteA4-4	KS	AT(H)	---	---	KR	ACP
PteA5	PteA5-1	KS	AT(H)	---	---	KR	ACP
	PteA5-2	KS	AT(CH3)	---	---	KR	ACP

Erläuterungen: KS* bezeichnet KS-Domänen des KSQ-Typs in Beladungsmodulen. AT(H) steht für AT-Domänen mit Substratspezifität für Malonyl-CoA, AT(CH3) für Methylmalonyl-CoA und AT(MB) für Methylbutyryl-CoA.

Die Größe der übrigen Cluster liegt zwischen 8 und 17,5 kbp. Deren Funktionalität und eventuelle Polyketid-Produkte waren in der Literatur noch nicht beschrieben. Von diesen kleineren Clustern zeigten auf Proteinebene nur die Sequenzen des Clusters Pks5 eine hohe Ähnlichkeit mit den Proteinen der drei großen Cluster. Um in den nachfolgenden phylogenetischen Analysen stabile Ergebnisse erzielen zu können, wurde die kleineren Cluster, abgesehen von Pks5, in die Datensätze der Domänen-Alignments nicht mit einbezogen.

3.3.2 Phylogenetische Einordnung in den Zusammenhang mit anderen PKS-I-Clustern aus Streptomyceten

Um die PKS-I-Cluster aus *S. avermitilis* zunächst in einen phylogenetischen Zusammenhang mit Clustern desselben Typs aus anderen Streptomyceten-Stämmen zu bringen, wurden die KS-Domänen von *S. avermitilis* in einen Datensatz aus KS-Domänen von 17 ausgewählten charakterisierten PKS-Biosynthesewegen von Streptomyceten integriert und das resultierende Aminosäure-Sequenz-Alignment für eine phylogenetische Analyse herangezogen. Die 17 charakterisierten PKS-I-Cluster sind in **Tab. 9** zusammengestellt. Die Baum-Rekonstruktion wurde mit Hilfe des Bayes-Verfahrens und der Distanz-basierten NJ-Methode durchgeführt. Das Ergebnis der Berechnung mit dem Bayes-Verfahren ist in **Abb. 18** gezeigt.

Die phylogenetische Rekonstruktion ergab, dass die Mehrzahl der Domänen in Cluster-spezifische Gruppen fällt, wobei die Hauptverzweigungen durch hohe *Clade probability values* sehr gut unterstützt wurden. Die Berechnungen mit der NJ-Methode führten zu der gleichen Cluster-spezifischen Topologie, wenngleich es geringfügige Unterschiede in der relativen Anordnung dieser Gruppen zueinander gab. Die KS-Domänen des *ave*- und des *pte*-Clusters bilden jeweils eine homogene Gruppe mit einer einzigen Ausnahme hinsichtlich des letzteren Clusters. Dies deutet darauf hin, dass die weit überwiegende Mehrzahl dieser KS-Domänen aus wiederholten Genduplikationen entstanden ist. Die meisten Sequenzen aus dem *olm*-Cluster liegen ebenfalls innerhalb einer *olm*-spezifischen phylogenetischen Gruppe, allerdings liegen insgesamt sechs KS-Domänen dieses Clusters außerhalb und damit in engerer Nachbarschaft zu Sequenzen anderer PKS-I-Cluster aus Streptomyceten.

Tab. 9: Ausgewählte PKS-I-Cluster aus Streptomycceten

Organismus	Polyketid-Verbindung	Proteine
<i>Amycolatopsis mediterranei</i>	Rifamycin	RifA, RifB, RifC, RifD, RifE
<i>Micromonospora griseorubida</i>	Mycinamicin	MycAI - MycAV
<i>Micromonospora megalomicea</i>	Megalomicin	MegAI, MegAII, MegAIII
<i>Saccharopolyspora spinosa</i>	Spinosad	SpnA - SpnE
<i>Saccharopolyspora erythrea</i>	Erythromycin	DEBS1, DEBS2, DEBS3
<i>Streptomyces antibioticus</i>	Oleanomycin	OleAI, OleAII
<i>Streptomyces caelestis</i>	Niddamycin	NidA1 – NidA5
<i>Streptomyces cinnamonensis</i>	Monensin	MonA1 – MonA8
<i>Streptomyces fradiae</i>	Tylacton	TyIG1 – TyIG5
<i>Streptomyces hygroscopicus</i>	Rapamycin	RapA, RapB, RapC
<i>Streptomyces hygroscopicus subsp. ascomyceticus</i>	FK-520 / Ascomycin	FkbA, FkbB, FkbC
<i>Streptomyces nanchangensis</i>	Nanchangmycin	NanA1 - NanA11
<i>Streptomyces natalensis</i>	Pimaricin	PimS0 – PimS4
<i>Streptomyces nodosus</i>	Amphotericin	AmphA - AmphK
<i>Streptomyces noursei</i>	Nystatin	NysA - NysK
<i>Streptomyces sp. FR-008</i>	Candididin	FscA - FscF
<i>Streptomyces venezuelae</i>	Pikromycin	PikAI - PikAIV

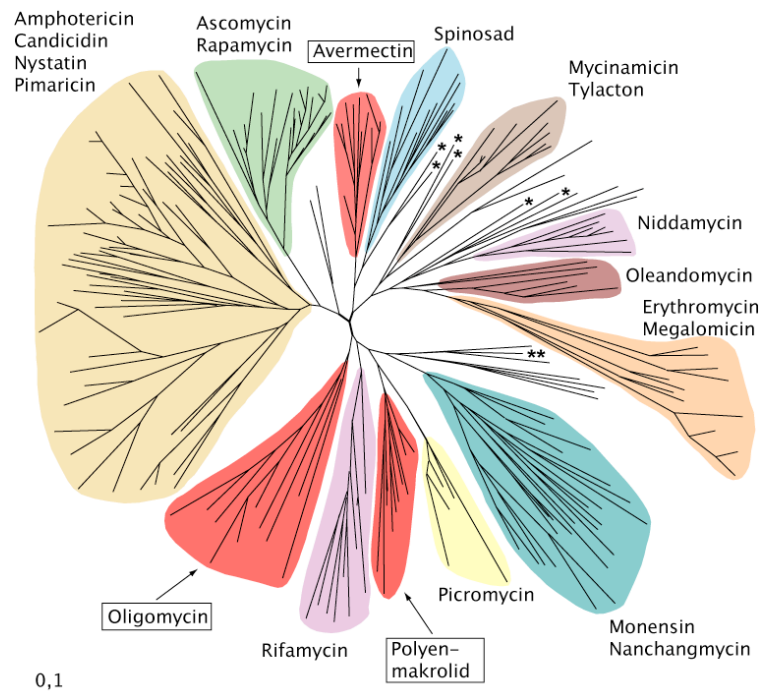


Abb. 18: Phylogenie der KS-Domänen von ausgewählten PKS-I-Clustern aus Streptomycceten.

Legende zu Abb. 18:

Der Baum wurde mit dem Bayes-Verfahren aus einem Aminosäure-Sequenz-Alignment berechnet. Die zu den drei großen PKS-I-Clustern aus *S. avermitilis* gehörenden KS-Domänen sind rot unterlegt. Außerhalb des Hauptcladus von Oligomycin und des Polyenmakrolids gelegene KS-Domänen sind mit einzelnen Sternen bzw. einem Doppelstern markiert.

3.3.3 Phylogenien der verschiedenen Domänen-Typen

Zum Zweck der Erstellung von Einzelphylogenien wurden aus den Nucleotid-Sequenzen der codierenden Gene und aus den Aminosäure-Sequenzen der Proteine der drei großen Cluster (*ave*, *pte* und *olm*) sowie des *pks5*-Clusters die Domänen ausgeschnitten und in Nucleotid-Alignments bzw. Aminosäure-Alignments zusammengestellt. Zur Berechnung der Bäume wurden zwei unterschiedliche Verfahren benutzt, zum einen das Bayes-Verfahren und zur ergänzenden Absicherung das MP-Verfahren. Für die Rekonstruktion der phylogenetischen Bäume mittels des Bayes-Verfahrens wurden diese Sequenz-Alignments in einen gemischten Datensatz gebracht und gemeinsam verrechnet, wobei die Nucleotid-Sequenzen in die drei Codon-Positionen aufgeteilt wurden. Für die alternative Berechnung mit dem MP-Verfahren wurden hingegen nur die Nucleotid-Alignments verwendet. Die Domänen der Beladungs-Module wurden nicht in die Analyse mit einbezogen. Es war möglich, stabile Phylogenien für die Domänen-Typen KS, AT, DH und KR zu erhalten. Eine Phylogenie von ER-Domänen konnte nicht erstellt werden, da im Datensatz nur drei ER-Domänen vorhanden waren. Die Berechnung eines phylogenetischen Baums der ACP-Domänen mit ausreichender Stabilität war aufgrund der Kürze dieser Sequenzen und ihre hohen Ähnlichkeit untereinander ebenfalls nicht möglich.

Die Bäume, die mit den beiden Methoden erhalten wurden, sind vergleichend in **Abb. 19** und **Abb. 20** gegenüber gestellt. Im Vergleich der Rekonstruktionsmethoden zeigte sich ganz überwiegend eine sehr gute Übereinstimmung, insbesondere im Hinblick auf die Identifizierung von Hauptverzweigungen innerhalb der Bäume. Die mit dem Bayes-Verfahren berechneten Phylogenien wiesen bei allen Abzweigungen von Hauptgruppen gute bis sehr gute *Clade probability values* auf, so dass von ausreichend stabilen und durch unterschiedliche Verfahren reproduzierbaren Phylogenien ausgegangen werden konnte. Um die Beziehungen zwischen den PKS-I-Clustern in eine Gesamtschau zu bringen, wurden die mit dem Bayes-Verfahren berechneten Einzelbäume in ein Cluster-Aufbauschema projiziert (**Abb. 21**).

Die phylogenetischen Bäume zeigen charakteristische Beziehungen in Abhängigkeit des Domänen-Typs. So besteht der AT-Baum aus zwei Hauptgruppen, wie sie bereits in der Analyse der AT-Domänen der PKS-I aus vollständig sequenzierten Bakteriengenomen zu Tage traten (siehe Abschnitt 3.2.3.2). Der eine Cladus umfasst diejenigen AT-Domänen, die Malonyl-CoA als Substrat benutzen, der andere solche, die Methylmalonyl-CoA verwenden. Der Baum der KR-Domänen setzt sich ebenfalls aus Hauptgruppen zusammen, was der funktionellen Einteilung gemäß der Stereochemie der resultierenden Hydroxyl-Gruppen der Polyketid-Kette entspricht, die auf der Grundlage von Sequenz-Vergleichen beschrieben wurde (Caffrey 2003) (siehe hierzu Abschnitt 1.2.2.1.1). Zusätzlich tritt einer dritter Typ auf, der insofern einen Mischtyp darstellt, als er die für die anderen beiden Typen kennzeichnenden Sequenzmotive in sich vereint, d. h. diese KR-Domänen besitzen sowohl das LDD-Motiv als auch den charakteristischen Tryptophan-Rest.

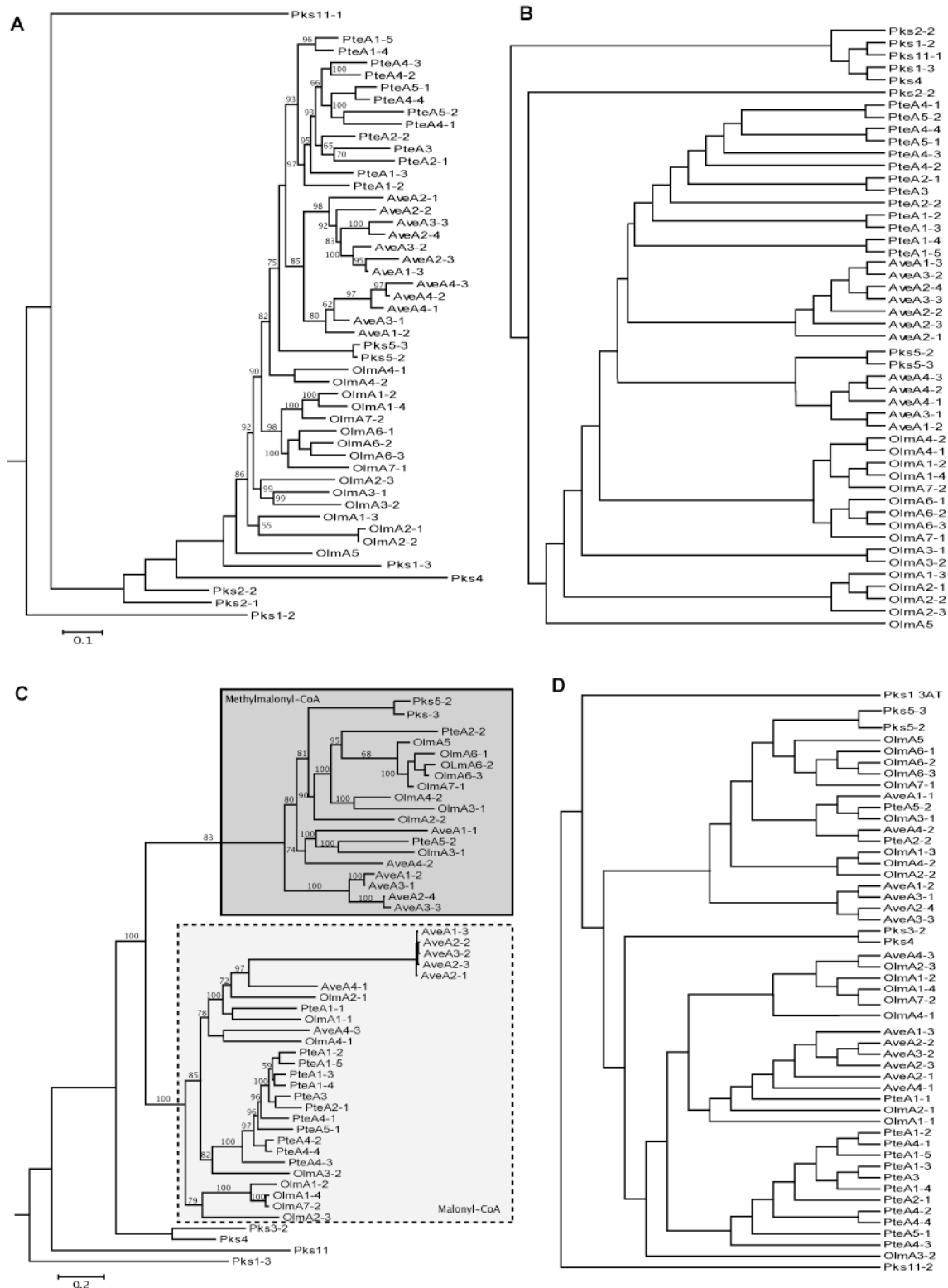


Abb. 19: Phylogenien der Domänen-Typen der PKS-I aus *S. avermitilis* (Teil 1).

(A) KS-Domänen gemäß Bayes-Verfahren. (B) KS-Domänen gemäß MP-Verfahren. (C) AT-Domänen gemäß Bayes-Verfahren. Die Malonyl-CoA-spezifischen und die Methylmalonyl-CoA-spezifischen Domänen sind durch unterschiedliche Grautöne hervorgehoben. (D) AT-Domänen gemäß MP-Verfahren.

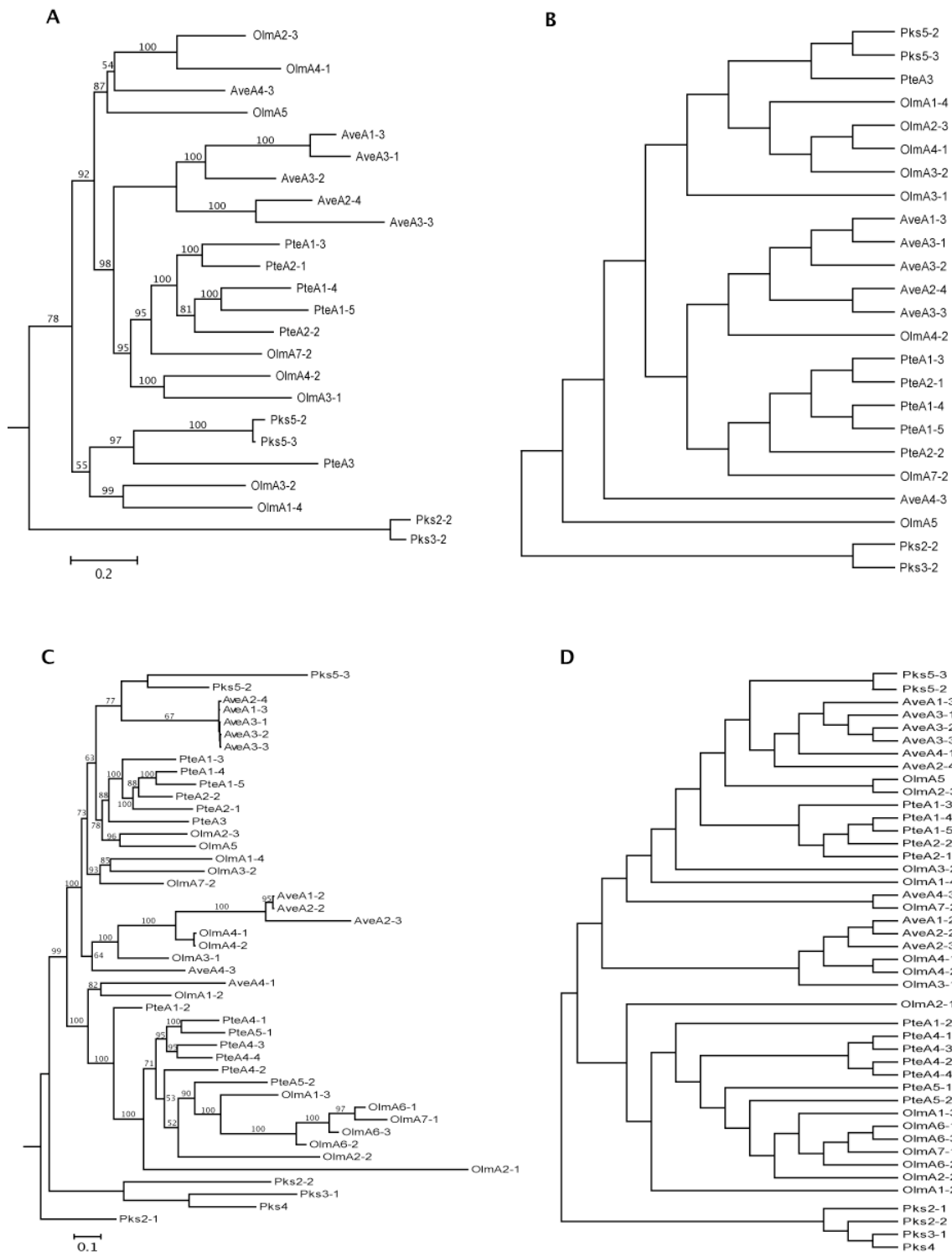


Abb. 20: Phylogenien der Domänen-Typen der PKS-I aus *S. avermitilis* (Teil 2).

(A) DH-Domänen gemäß Bayes-Verfahren. (B) DH-Domänen gemäß MP-Verfahren. (C) KR-Domänen gemäß Bayes-Verfahren. (D) KR-Domänen gemäß MP-Verfahren. Zahlen über den Zweigen geben die *Clade probability values* an. Bei den mit dem Bayes-Verfahren berechneten Bäumen wurde die *Midpoint rooting*-Methode angewendet.

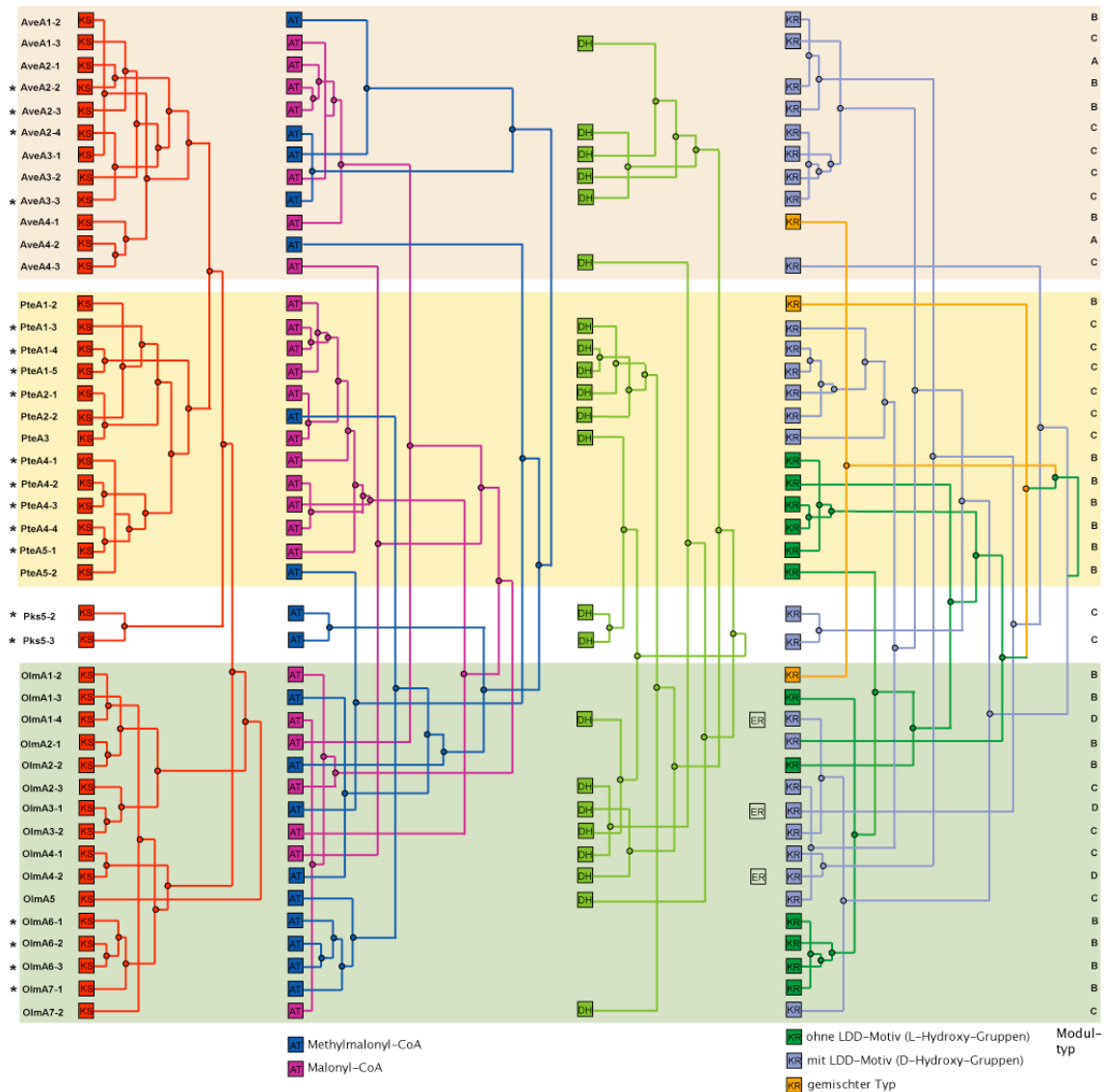


Abb. 21: Projektion der Phylogenien der verschiedenen Domänen-Typen auf das Cluster-Schema.

Vollständige Kongruenz zeigende Module sind links mit einem Stern markiert. Im Fall der AT-Domänen sind die unterschiedlichen Substratspezifitäten durch verschiedene Farben gekennzeichnet. Ebenso sind unterschiedlichen Arten von KR-Domänen farblich unterschieden. Die am rechten Rand stehende Klassifizierung in Modul-Typen bezieht sich auf die Domänen-Zusammensetzung (vgl. **Abb. 5** in Abschnitt 1.2.2.1.1 der Einleitung).

Die Domänen-Einzelphylogenien ermöglichten, die Module im Hinblick auf ihre Kongruenz untereinander zu vergleichen. Dabei bedeutet Kongruenz, dass mindestens zwei Module in allen Domänen phylogenetisch übereinstimmen, also die Domänen des einen Moduls immer im selben Cladus wie die Domänen des anderen Moduls zu finden sind. Die systematische Auswertung ergab, dass sich insgesamt 15 Module auf diese Weise als kongruent klassifizieren ließen. Diese sind in **Abb. 21** durch Sterne markiert. Die Entstehung dieser Module kann durch einfache Genduplikationen ohne weitere

Veränderungen der Modul-Zusammensetzung interpretiert werden. Alle anderen Module, die einen Anteil von 65 % ausmachen, sind dagegen phylogenetisch inkongruent, d.h. zumindest eine Domäne passt nicht in den phylogenetischen Zusammenhang der anderen im Modul vorhandenen Domänen. Diese nicht passenden Komponenten treten auffälligerweise nicht gleichermaßen in allen Domänen-Typen auf. Wie bereits aus der Gesamtphylogenie von KS-Domänen aus Streptomyceten-PKS-I erkennbar war (**Abb. 18**), liegen die KS-Domänen der Cluster von *S. avermitilis* praktisch alle innerhalb einer phylogenetischen Gruppe und lassen sich daher als Abkömmlinge einer Cluster-spezifischen gemeinsamen Vorläufer-Sequenz interpretieren. Sie sind aus diesem Vorläufer durch wiederholte Genduplikationen hervorgegangen, ohne dass es zu einer Vermischung zwischen den Clustern kam. Im Gegensatz dazu fanden sich innerhalb eines bestimmten Clusters solche AT-Domänen, die nahe gemeinsame Vorfahren mit AT-Domänen eines anderen Clusters besitzen. Beispielsweise ist die AT-Domäne des Moduls AveA4-2 phylogenetisch näher verwandt mit denen der Module OlmA3-1 und PteA5-2 als mit den anderen Methylmalonyl-CoA-spezifischen AT-Domänen des ave-Clusters. Dasselbe gilt für die DH- und KR-Domänen, wenngleich in allen Fällen die Mehrzahl der Module von einem Cluster-eigenen gemeinsamen Vorläufer abstammt. Für die in Modulen des C-Typs (KS-AT-DH-KR-ACP) auftretenden DH-KR-Kombinationen ließ sich bestätigen, dass DH-Domänen immer mit solchen KR-Domänen einhergehen, die das LDD-Motiv besitzen und folglich bei der Reduktion der Keto-Gruppe das D-Isomer erzeugen. Auch sind die meisten DH-KR-Gruppierungen phylogenetisch kongruent, was durch einen Vergleich der entsprechenden zwischen den DH- und KR-Domänen befindlichen Interdomänen-Bereiche zusätzlich bestätigt wird. Diese zeigen nämlich im Allgemeinen eine sehr deutlich ausgeprägte Sequenz-Ähnlichkeit (Daten nicht gezeigt). Hiervon gibt es nur einige wenige Ausnahmen, z.B. die Module PteA3 und OlmA2-3.

Die aus den Einzelphylogenien der Domänen ableitbaren Inkongruenzen weisen auf Rekombinationsprozesse hin, die nach einer Genduplikation stattgefunden haben. Dabei fällt auf, dass diese Prozesse im Genom von *S. avermitilis* nur Domänen-Typen betreffen, die wie im Fall der AT- und KR-Domänen in enzymatisch unterschiedlichen Varianten vorkommen oder deren Anwesenheit die chemische Struktur der resultierenden Polyketid-Kette verändert, wie es auf die KR-, DH- und ER-Domänen zutrifft. In den nächsten Schritten der Untersuchung wurden aus dem Gesamtschema der **Abb. 21** solche Konstellationen herausgesucht und im Detail analysiert, aus denen sich die abgelaufenen Rekombinationsprozesse klar ableiten ließen.

3.3.4 Änderung der Substratspezifität durch Ersetzung von AT-Domänen

Aus dem Syntheseprinzip modularer PKS-I lässt sich vorhersagen, dass die Änderung der Carbonsäure-Einheit, die in die wachsende Polyketid-Kette eingebaut werden soll, eine wichtige Möglichkeit zur Generierung von Variabilität darstellt. In den PKS-I-Modulen aus *S. avermitilis* konnten Beispiele für derartige AT-Domänen-Ersetzungen identifiziert werden. Die Module PteA1-3, PteA1-4, PteA2-1 und PteA2-2 gehören alle zum C-Typ, besitzen also neben den Domänen der Basiskonfiguration zusätzlich DH- und KR-Domänen. Die KS-, DH- und KR-Domänen jedes dieser Module sind phylogenetisch kongruent, d.h. sie liegen in denselben Untergruppen der jeweiligen Bäume (siehe **Abb. 22**). Die AT-Domänen zeigen ebenfalls Kongruenz, jedoch mit Ausnahme der AT-Domäne von PteA2-2. Diese ist die einzige innerhalb dieser Modul-Gruppe, die als Substrat Methylmalonyl-CoA verwendet. Die aus

phylogenetischer Sicht nächsten Nachbarn sind die AT-Domänen von OlmA5 und der Module von OlmA6. Die Interdomänen-Bereiche, die vor der AT-Domäne liegen weisen über ihre gesamte Länge ausgeprägte Sequenzhomologie auf Nucleotid-Ebene auf. Ebenso finden sich solche ausgeprägt homologen Bereiche in der hinter der AT-Domäne befindlichen Interdomänen-Region. Auffällig ist, dass sich die AT-DH-Interdomänen-Region von PteA2-2 als Hybrid-Sequenz darstellt: Im 5'-Bereich ist sie den Sequenzen des *olm*-Clusters ähnlicher, im näher an der DH-Domäne liegenden Bereich hingegen zeigt sich eine größere Verwandtschaft mit den Sequenzen aus dem *pte*-Cluster. Das spricht für Rekombinations-Bruchstellen innerhalb der Interdomänen-Bereich vor und hinter der betroffenen AT-Domäne.

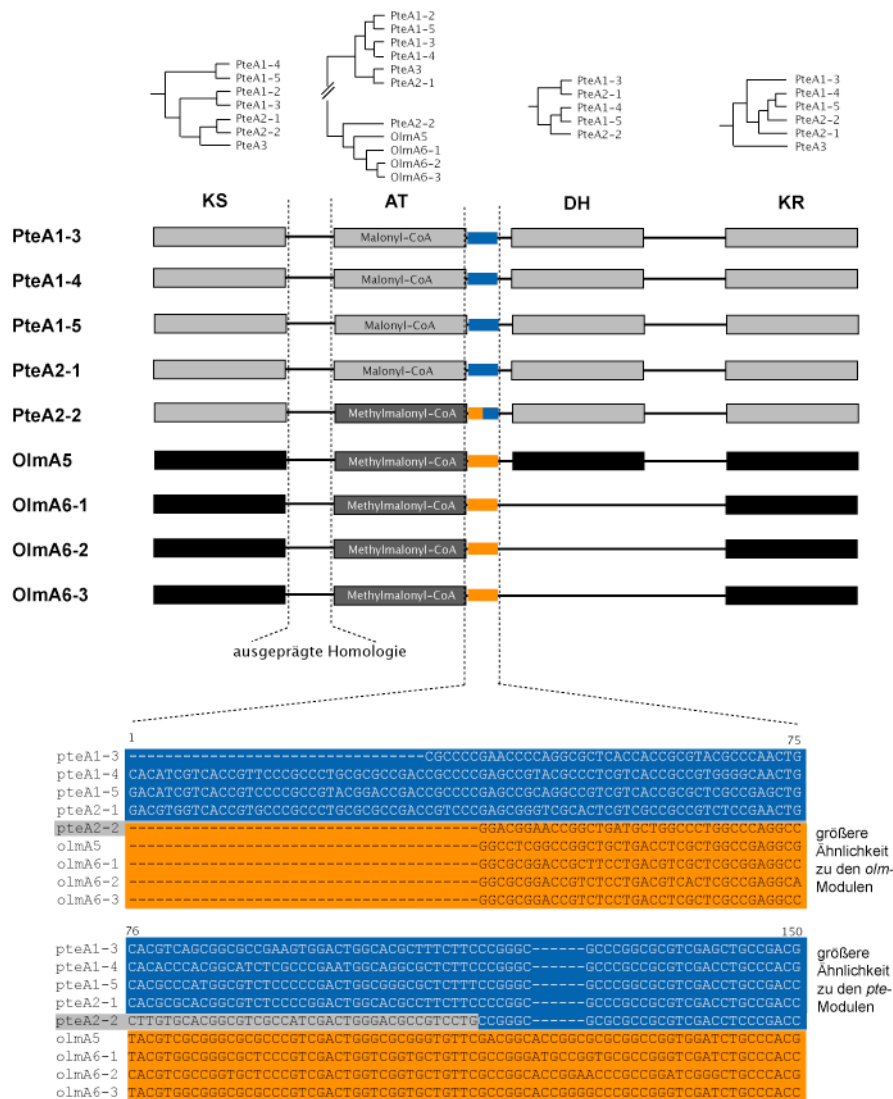


Abb. 22: Ersetzung einer AT-Domäne durch Rekombination zwischen Modulen.

Die phylogenetische Inkongruenz der AT-Domänen ist anhand der entsprechenden Baumausschnitte gezeigt. Die AT-DH-Interdomänen-Bereiche sind in blau und orange hervorgehoben, um den hybriden Charakter des Interdomänen-Bereichs des Moduls PteA2-2 zu verdeutlichen. Nähere Erläuterungen werden im Text gegeben.

Eine sehr ähnliche Konstellation konnte in Modulen des *ave*-Clusters identifiziert werden. Hier betrifft es die Modul-Gruppe AveA1-3, AveA3-2, AveA2-4 und AveA3-3, von denen die ersten beiden AT-Domänen mit Spezifität für Malonyl-CoA besitzen, die letzteren beiden dagegen solche mit Spezifität für Methylmalonyl-CoA.

3.3.5 Veränderungen der Reduktionsstufe bewirkende Domänen-Ersetzungen

Um die theoretisch durch Rekombinationsprozesse möglichen Veränderungen der Reduktionsstufe der Polyketid-Kette festzustellen, wurden die Nucleotid-Sequenzen aller Modul-Typen zunächst auf Homologie-Muster hin untersucht. Die Ergebnisse sind schematisch in **Abb. 23** zusammengestellt.

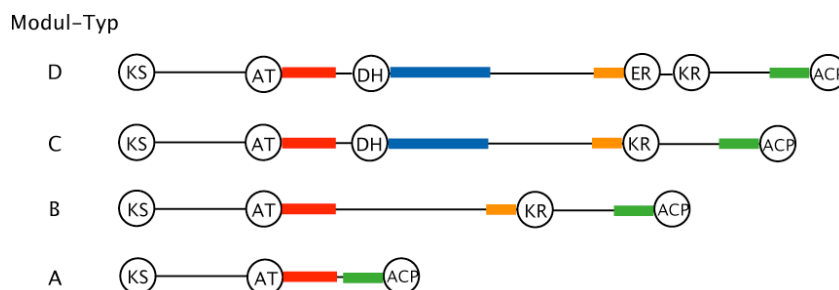


Abb. 23: Homologe Abschnitte in der Nucleotid-Sequenz der unterschiedlichen Modul-Typen.

Die homologen Bereiche sind farblich gegeneinander abgehoben.

Sequenz-Homologie konnte in allen KR-ACP-Interdomänen-Bereichen gefunden werden. Der die letzten etwa 150 Nucleotide umfassende Abschnitt ist auch Teil der AT-ACP-Interdomänen-Bereiche des Modul-Typs A (in **Abb. 23** grün eingetragen). Die homologen Sequenz-Bereiche zwischen AT- und DH-Domänen finden sich auch am Anfang der Interdomänen-Bereiche, die bei Modulen des Typs B zwischen AT- und KR-Domänen liegen und bei Modulen des Typs A zwischen AT- und ACP-Domänen lokalisiert sind (in **Abb. 23** rot eingetragen). Die Interdomänen-Abschnitte zwischen DH- und KR-Domänen einerseits und diejenigen zwischen AT- und KR-Domänen andererseits weisen zwar typischerweise eine sehr ähnliche Länge von etwa 300 Aminosäure-Resten auf, sind jedoch im Anfangsbereich nicht homolog. Der ca. 400 bp umfassende Anfangsabschnitt, der bei Modulen des Typs D zwischen DH – und ER-Domänen und bei Modulen des Typs C zwischen DH- und KR-Domänen liegt (in **Abb. 23** blau eingetragen), ist nicht Bestandteil des entsprechenden Abschnitts in Modulen des Typs B. Ein weiterer homologer Sequenz-Bereich findet sich wiederum einheitlich immer direkt vor KR- und ER-Domänen (in **Abb. 23** orange eingezeichnet). Im Folgenden wurde untersucht, welche Arten von Domänen-Austauschen oder Ersetzung von Domänen-Gruppen durch andere in den PKS-I-Modulen von *S. avermitilis* zu beobachten sind.

Die Umwandlung des Modul-Typs A in den Typ B oder umgekehrt konnte anhand des Moduls AveA2-1 gezeigt werden. Dieses ist eines der beiden im Datensatz vorhandenen Module des Typs A (vgl.

Tab. 8). Seine KS- und AT-Domäne stehen in jeweils enger Beziehung zu den entsprechenden Domänen des Moduls AveA2-2, das über eine KR-Domäne verfügt und somit zum Modul-Typ gehört (**Abb. 24**).

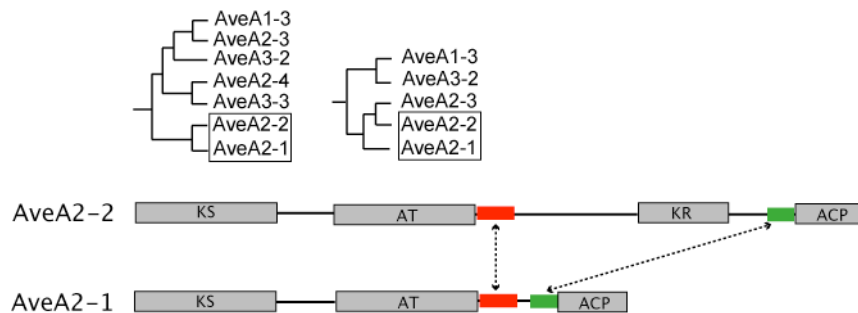


Abb. 24: Umwandlung zwischen den Modul-Typen A und B.

Der Vergleich der AT-ACP- bzw. AT-KR-Interdomänen-Bereiche ergab, dass sie im Anfangsbereich ein Segment von etwa 150 bp Länge besitzen, das in dieser Form nur in Modulen des *ave*-Clusters vorkommt. Außerdem weisen beide Modul-Sequenzen den generell vorhandenen Homologie-Bereich direkt vor der ACP-Domäne auf. Die vorliegende Konstellation lässt sich schlüssig auf folgende Weise deuten: Entweder bildete sich AveA2-1 infolge des Verlustes eines großen Sequenz-Bereiches zwischen AT- und ACP-Domäne unter Einschluss der KR-Domäne, entsprechend einer Umwandlung von Typ B in Typ A, oder aus einem ursprünglichen Typ A-Modul, wie es AveA2-1 darstellt, entwickelte sich das Typ B-Modul durch Integration einer KR-Domäne zusammen mit der charakteristischen langen Interdomänen-Region. Bei beiden Alternativen wären die Rekombinationsstellen in den jeweiligen, die ausgetauschten Sequenz-Elemente umgebenden Interdomänen-Bereichen lokalisiert.

Darüber hinaus konnten Beispiele für den vollständigen Transfer einer DH-KR-Einheit gefunden werden. Stellt man die vier Module AveA1-3, AveA3-2, AveA2-2 und AveA2-3 einander gegenüber, so fällt auf, dass die ersteren beiden Module des Typs C sind, die anderen beiden hingegen zum Modul-Typ B gehören (siehe **Abb. 25**). Dennoch zeigen sowohl ihre KS- als auch AT-Domänen vollständige phylogenetische Kongruenz. Abgesehen von der Existenz von DH-Domänen in AveA1-3 und AveA3-2, die in den beiden anderen Modulen fehlen, zeigte sich ein weiterer Unterschied. Die KR-Domänen der Modul-Paare gehören nämlich zu verschiedenen Typen. Während die mit DH-Domänen ausgestatteten Module erwartungsgemäß KR-Domänen mit LDD-Motiv aufweisen, besitzen die zwei Module des C-Typs den anderen KR-Domänen-Typ. Die Beziehungen der Module zueinander lässt sich am einfachsten damit erklären, dass eine DH-KR-Einheit gegen eine nur eine KR-Domäne aufweisende Sequenz-Einheit ersetzt wurde. Diese Konversion wird zusätzlich durch die sehr ausgeprägte Homologie der AT-DH-Interdomänen-Bereiche einerseits und der ersten etwa 200 bp zwischen AT- und KR-Domäne andererseits gestützt.

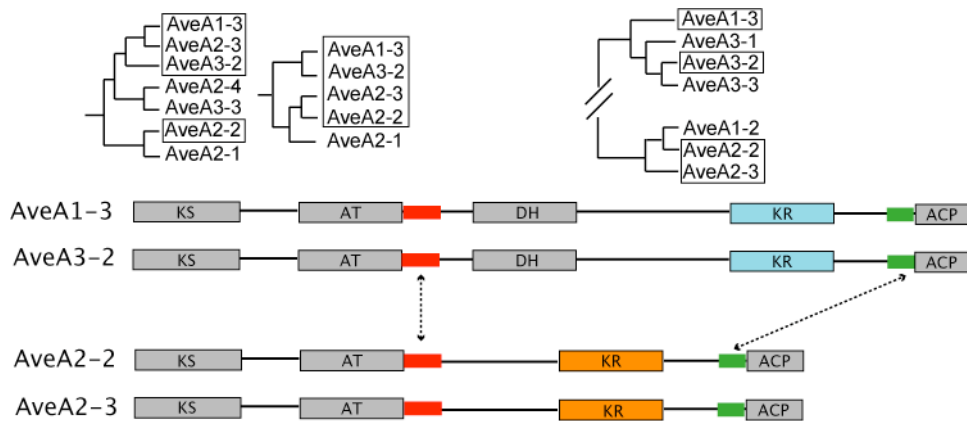


Abb. 25: Transfer einer DH-KR-Einheit zwischen Modulen.

Die roten und grünen Balken markieren homologe Bereiche in der Nucleotid-Sequenz entsprechend zu **Abb. 23**. Die beiden verschiedenen Typen an KR-Domänen sind durch unterschiedliche Farben (blau für den Typ mit LDD-Motiv, orange für den Typ ohne dieses Motiv) hervorgehoben.

Auf ähnliche Weise könnte das zweite Modul des Proteins PteA1 seine DH-Domäne durch Rekombination mit der Sequenz eines Moduls von PteA4 verloren haben (**Abb. 26**). Diese Annahme wird unterstützt durch den Umstand, dass die KR-Domäne von PteA1-2 nicht zum selben Typ wie die anderen KR-Domänen innerhalb des Proteins PteA1 gehört. Statt dessen zeigt es auf der Aminosäure-Sequenz-Ebene im Anfangs- und Endbereich eine sehr hohe Ähnlichkeit mit den KR-Domänen der PteA4-Module, und zwar 92 % bezogen auf die ersten 80 Aminosäure-Reste und 91 % bezogen auf letzten 50 Reste.

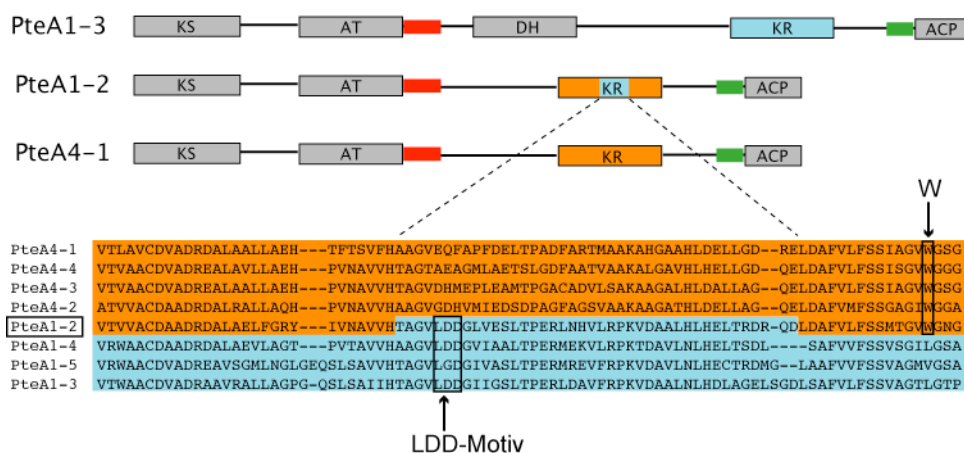


Abb. 26: Ausbildung einer KR-Mischdomäne.

KR-Domänen mit LDD-Motiv sind blau dargestellt, die des anderen Typs, der einen charakteristischen Tryptophan-Rest aufweist, in orange. Das Aminosäure-Sequenz-Alignment zeigt, dass es sich bei der KR-Domäne des Moduls PteA1-2 um einen Mischtyp handelt, der Sequenz-Charakteristika beider Typen von KR-Domänen in sich vereint.

Der Sequenz-Abschnitt dazwischen scheint von dem anderen KR-Domänen-Typ zu stammen, da eine Ähnlichkeit von 87 % mit den entsprechenden Abschnitten der anderen KR-Domänen von PteA1 ermittelt wurde. Wie diese Domänen, weist die KR-Domäne von PteA1-2 das LDD-Motiv auf.

Wahrscheinlich wurde die in einem Vorläufer-Modul ursprünglich vorhandene DH-KR-Einheit mittels Rekombination durch eine Sequenz ersetzt, die nur für eine KR-Domäne codierte, wobei es gleichzeitig zu einer Änderung des KR-Domänen-Typs kam. Ein zweites Rekombinationsereignis innerhalb der KR-Domäne formte dann den zentralen Sequenz-Abschnitt der Domäne in den ursprünglichen Typ zurück. Anhand dieser Sequenz-Kombination konnte gezeigt werden, dass die Grenzen der rekombinierten Abschnitte nicht grundsätzlich in den Interdomänen-Bereichen liegen.

Inwieweit ER-Domänen an solchen Austauschprozessen beteiligt sein könnten, ließ sich im Rahmen des benutzten Datensatzes nur schwer ermitteln, da er nur drei Module aufweist, die zum Typ D gehören. Diese Module, OlmA1-4, OlmA3-1 und OlmA4-2, gehören alle zum *olm*-Cluster (vgl. **Tab. 8**). In Bezug auf diese Module ließen sich folgende Daten gewinnen. Die ER-Domäne von OlmA1-4 zeigt auf der Ebene der Aminosäure-Sequenz eine Ähnlichkeit von 72 % mit derjenigen von OlmA3-1 und von 75 % mit der von OlmA4-2. Die beiden letztgenannten Module sind phylogenetisch kongruent im Hinblick auf ihre DH- und KR-Domänen, ebenso ist der Interdomänen-Bereich, der die DH- mit der ER-Domäne verbindet, mit jeweils ca. 315 Aminosäure-Resten fast gleich lang und sehr ähnlich zusammengesetzt. Die Ähnlichkeit der ER-Domänen von OlmA3-1 und OlmA4-2 beläuft sich auf 88 %. Im Gegensatz hierzu fallen die DH- und die KR-Domäne des Moduls OlmA1-4 in andere phylogenetische Gruppen, und der Bereich zwischen DH- und ER-Domäne fällt mit 345 Aminosäure-Resten deutlich länger als bei den anderen beiden Modulen des Typs D aus, was auf mehrere Insertionen zurückzuführen ist. In den Nucleotid-Sequenzen war eine ausgeprägte Homologie innerhalb eines Segmentes von etwa 250 bp direkt vor der ER-Domäne erkennbar. Außerdem fand sich am Anfang der KR-Domänen ein Sequenz-Motiv, das nur in den Modulen des Typs D vorzukommen scheint. Zusammengefasst ließe sich der derzeitige Aufbau der Module durch folgende Prozesse erklären. Zunächst wurde eine ER-Domäne von dem Vorläufer des Moduls OlmA4-2 auf den zu diesem Zeitpunkt noch keine ER-Domäne aufweisenden Vorläufer von OlmA1-4 übertragen, wodurch dieser in einen D-Typ überführt wurde.

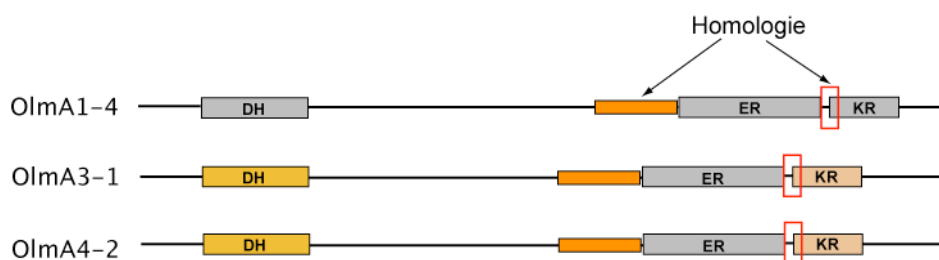


Abb. 27: Sequenz-Ähnlichkeiten in den ER-Domänen-enthaltenden Modulen des Typs D.

Der Interdomänen-Abschnitt des Moduls OlmA1-4 ist ca. 10 % länger als derjenige der anderen beiden Module des Typs D. Deren Übereinstimmungen betreffen auch die phylogenetische Kongruenz der DH- und KR-Domänen, die durch Farben hervorgehoben ist. Der rote Kasten markiert einen Sequenz-Bereich, der nur in Modulen des Typs D vorkommt.

Dabei waren die Rekombinationsbruchstellen zum einen im Endbereich der Interdomänen-Regionen zwischen DH und ER bzw. DH und KR und zum anderen im Anfangsbereich der KR-Domänen lokalisiert. Wie in den weiter oben beschriebenen Beispielen wäre auch der umgekehrte Prozess denkbar. Später könnte dann das gesamte DH-ER-KR-Segment auf den Vorläufer von OlmA3-1 übertragen worden sein.

Insgesamt betrachtet, blieb die Situation in Bezug auf ER-Domänen vergleichsweise unklar. Es konnten jedoch Hinweise darauf gefunden werden, dass sowohl der Austausch einzelner ER-Domänen als auch der Transfer großer Sequenzeinheiten, die eine DH-, ER- und KR-Domäne umfassen, möglich sein könnte.

3.3.6 Selektion in den PKS-I-Clustern von *S. avermitilis*

Um die im Laufe der evolutionären Entwicklung stattgefundenen Selektionsprozesse beurteilen zu können, wurde in den Nucleotid-Sequenzen der PKS-I-Gene die synonymen (p_s) und nicht-synonymen (p_n) Basenaustausche je Position bestimmt. Synonyme Austausche sind solche, die zu einem für dieselbe Aminosäure wie im Ausgangszustand codierenden Codon führen, während ein nicht-synonymer Basenaustausch die Veränderung der Aminosäure an der entsprechenden Position des Proteins zur Folge hat. Die Ergebnisse sind in **Tab. 10** zusammengestellt.

Tab. 10: Über die Sequenz gemittelten durchschnittlichen Basenaustausche in den verschiedenen Domänen-Typen der PKS-I-Gene von *Streptomyces avermitilis*

Domäne	p_s	S.F. (p_s)	p_n	S.F. (p_n)	p_n/p_s	Negative Selektion	Positive Selektion
KS	0,853	0,025	0,213	0,012	0,25	P=0,00000 Z=3,7	P=1,0000 Z=3,5
AT (H)	0,847	0,033	0,385	0,022	0,45	P=0,00000 Z=1,0	P=1,0000 Z=1,43
AT (C)	0,827	0,038	0,365	0,021	0,44	P=0,00000 Z=0,65	P=1,0000 Z=0,76
DH	0,980	0,038	0,478	0,0317	0,49	P=0,00000 Z=0,7	P=1,0000 Z=1,35
KR (A)	0,780	0,050	0,307	0,025	0,40	P=0,00000 Z=8,4	P=1,0000 Z=8,6
KR (B)	0,851	0,040	0,417	0,029	0,49	P=0,00000 Z=9,2	P=1,0000 Z=9,3

p_s : Zahl der synonymen Austausche je Position, p_n : Zahl der nicht-synonymen Austausche je Position, S.F.: Standardfehler, P: Wahrscheinlichkeit, Z: Z-Wert der Z-Statistik

Der kleinste Wert für das Verhältnis p_N/p_S wurde für die KS-Domänen gefunden. Für die anderen Domänen-Typen lagen die Werte zwischen 0,40 und 0,49. In allen Fällen war der Wert von p_S significant größer als der von p_N . Wie an den P-Werten zu erkennen ist, liegt bei allen Domänen-Typen ausschließlich negative Selektion vor. Die Hypothese auf positive Selektion wurde hingegen in untersuchten Fällen zurückgewiesen. Die gefundenen Werte stehen in guter Übereinstimmung mit diesbezüglichen Daten, die für bakterielle Paraloge erhältlich sind (Lynch und Conery 2000). Da bei der Betrachtung von Durchschnittswerten über längere Sequenz-Abschnitte eventuell Selektionseffekte, die auf kürzere Bereiche beschränkt sind, übersehen werden können, wurden auch *Sliding window*-Analysen über die kompletten Datensätze der einzelnen Domänen-Typen durchgeführt. Allerdings konnten auch mit diesem Verfahren keine Bereiche unter positiver Selektion identifiziert werden. Insgesamt sprachen daher die Ergebnisse eindeutig dafür, dass in den Genen der PKS-I von *S. avermitilis* negative Selektion zur Funktionserhaltung vorherrscht, während keine Anzeichen von Mutationen, die einer positiven Selektion unterlägen, zu finden waren.

3.4 Untersuchung der Genom-Sequenz von *Nostoc punctiforme* auf PKS- und NRPS-Cluster

Als Ausgangspunkt der phylogenetischen Analysen, die in den Abschnitten 3.2 und 3.3 dargestellt sind, wurden auch die Nucleotid- und Aminosäure-Sequenzen von PKS und NRPS in *N. punctiforme* durch Blast-Suchen identifiziert. Ergänzend hierzu wurden diese Sequenzen in den genomischen Zusammenhang eingeordnet, um den Aufbau der entsprechenden Gencluster zu bestimmen. Dies geschah mit Hilfe der Annotation, die auf der Website des *N. punctiforme*-Genomprojekts des „DOE Joint Genome Institute“ zur Verfügung gestellt wird. Da die dort angegebenen Annotationen noch vorläufigen Charakter haben, wurden die Sequenzen in unklaren Fällen manuell durch verschiedene Internet-basierte Homologie-Suchsysteme überprüft.

Das Genom von *N. punctiforme* enthält insgesamt 11 PKS- und NRPS-Cluster. Diese sind in **Tab. 11** aufgelistet. Die Größenangaben beziehen sich dabei auf Sequenzabschnitte, deren offene Leserahmen unter funktionellen Gesichtspunkten die Zuordnung zu einem gemeinsamen Synthese-System wahrscheinlich machen. In unmittelbarer Nähe zu PKS- und NRPS-Sequenzen gelegene Abschnitte, die für Transporter oder Transkriptionsregulatoren codieren, wurden in die Größenberechnung mit einbezogen. Da mit einer Ausnahme alle gefundenen Cluster zu Synthasen gehören, deren Produkte nicht bekannt sind, hat die Bestimmung von möglichen Cluster-Grenzen notwendigerweise vorläufigen Charakter.

Tab. 11: PKS- und NRPS-Cluster im Genom von *N. punctiforme*

Cluster-Nr.	Metabolit	Größe (kpb)	Komponenten
I	unbekannt	28,4	PKS (Glycolipidsynthesen
II	unbekannt	20,3	PKS (Glycolipidsynthesen)
III	unbekannt	32,3	PKS
IV	Nostopeptolid	41,2	NRPS und PKS
V	unbekannt	27,2	NRPS
VI	unbekannt	63,4	NRPS und PKS
VII	unbekannt	62,7	PKS und NRPS
VIII	unbekannt	11,8	PKS
IX	unbekannt	67,3	PKS und NRPS
X	unbekannt	24,5	PKS und NRPS
XI	unbekannt	24,8	NRPS und PKS

Tab. 11 zeigt, dass die Größen der PKS- und NRPS-Cluster in *N. punctiforme* zwischen ca. 20 und 70 kbp liegen. Vier der Cluster enthalten ausschließlich PKS-Komponenten (Nr. I, II, III und VIII), eines weist nur NRPS auf (Nr. V). Die übrigen sechs Cluster sind Mischformen, die beide Arten von Synthese-Systemen vereinen. In diese Gruppe fällt auch das einzige Cluster, dessen Produkt bekannt ist. Es handelt sich um die Nostopeptolid-Synthetase, die zuerst in dem Stamm *Nostoc* sp. GSV224 sequenziert und biochemisch charakterisiert wurde (Hoffmann *et al.* 2003), nachdem die Metabolite, Nostopeptolide A1, A2 und A3, aus demselben Stamm isoliert und ihre Strukturen aufgeklärt worden waren (Golakoti *et al.* 2000). Später konnten auch aus *N. punctiforme* die Nostopeptolide A1 und A3 isoliert werden (Hunsucker *et al.* 2004).

Im Folgenden werden die im Genom von *N. punctiforme* identifizierten Gencluster der Reihe nach schematisch gezeigt und beschrieben.

Das Cluster I (**Abb. 28**) codiert für eine Glycolipidsynthase. Neben den PKS-Sequenzen enthält es Glycosyltransferasen und die Komponenten eines ABC-Transporters. Eng assoziiert mit den Synthesen liegt ein ORF, der für eine PPTase codiert. Die Glycolipidsynthase NpR0043 zeigt die typischen multiplen ACP-Domänen, wie sie in anderen Bakterien beschrieben wurden. Anhand der Annotationen lässt sich auch Cluster II (**Abb. 29**) der Glycolipidsynthese zuordnen. Wie im voran gegangenen Fall ist ein ABC-Transporter Teil des Clusters. Wiederum ist eine PKS-Komponente (NpR1294) durch das mehrfache Auftreten von ACP-Domänen gekennzeichnet.

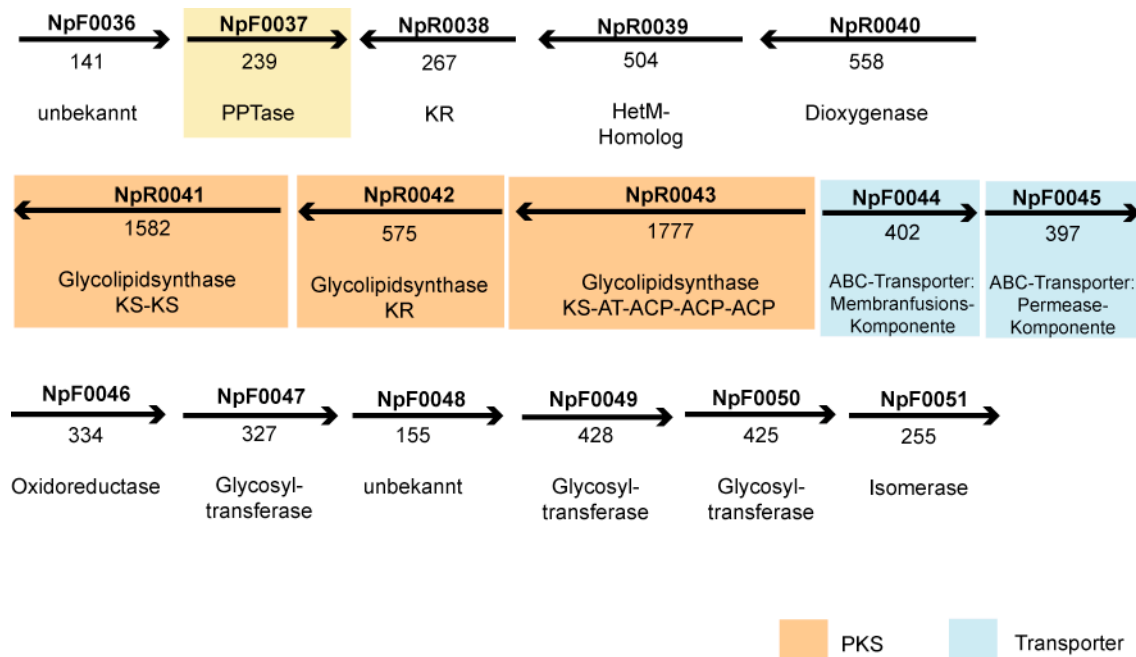


Abb. 28: Cluster I aus *N. punctiforme*.

Die Pfeilrichtungen beziehen sich auf die Orientierung der ORFs im Genom. Über den Pfeilen ist die Gennummer angegeben, direkt unter den Pfeilen die Länge der codierten Proteine in Aminosäure-Resten. Danach folgt die aus Sequenz-Homologien abgeleitete Annotation. Im Fall der PKS ist darüber hinaus auch die Domänen-Zusammensetzung verzeichnet. PKS und Transporter-Proteine sind farblich hervorgehoben. Eng mit dem Cluster assoziiert liegt ein ORF, der für eine PPTase codiert (gelb hervorgehoben).

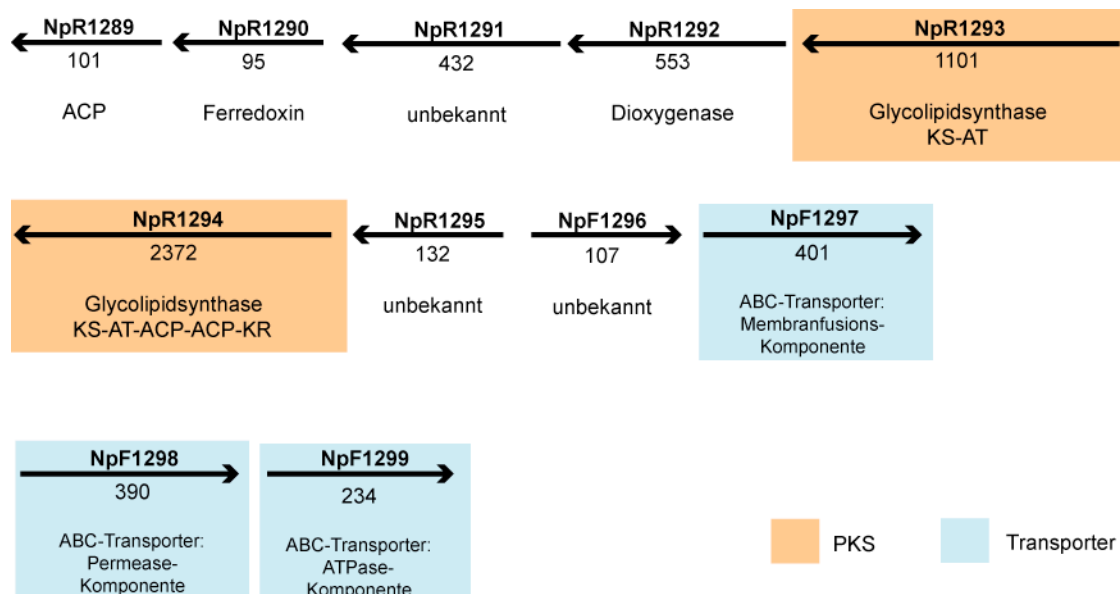


Abb. 29: Cluster II.

Darstellung und Bezeichnungen wie in Abb. 28.

Das Cluster III (**Abb. 30**) enthält vier jeweils einmodulare PKS-I-Proteine in ununterbrochener Abfolge. Wenige ORFs entfernt liegen die codierenden Sequenzen einer PPTase und eines Proteins, das Homologie zu Response-Regulatoren von Zwei-Komponenten-Systemen hat. Das Cluster umfasst ein FkbH-Homolog als Einzelprotein. FkbH-Homologe finden sich in vielen Polyketidsynthasen, z. B. im FK520-Gencluster von *Streptomyces hygroscopicus*, wo es möglicherweise an der Synthese ungewöhnlicher Polyketidketten-Bausteine wie Methoxymalonyl-CoA beteiligt ist (Wu *et al.* 2000a). Eine TE-Domäne ließ sich nicht sicher detektieren. Allerdings weist NpR2077, eines der als unbekannt klassifizierten Proteine, teilweise Sequenzähnlichkeiten mit TE-Domänen auf. Stromaufwärts sind Transposase-Gene lokalisiert.

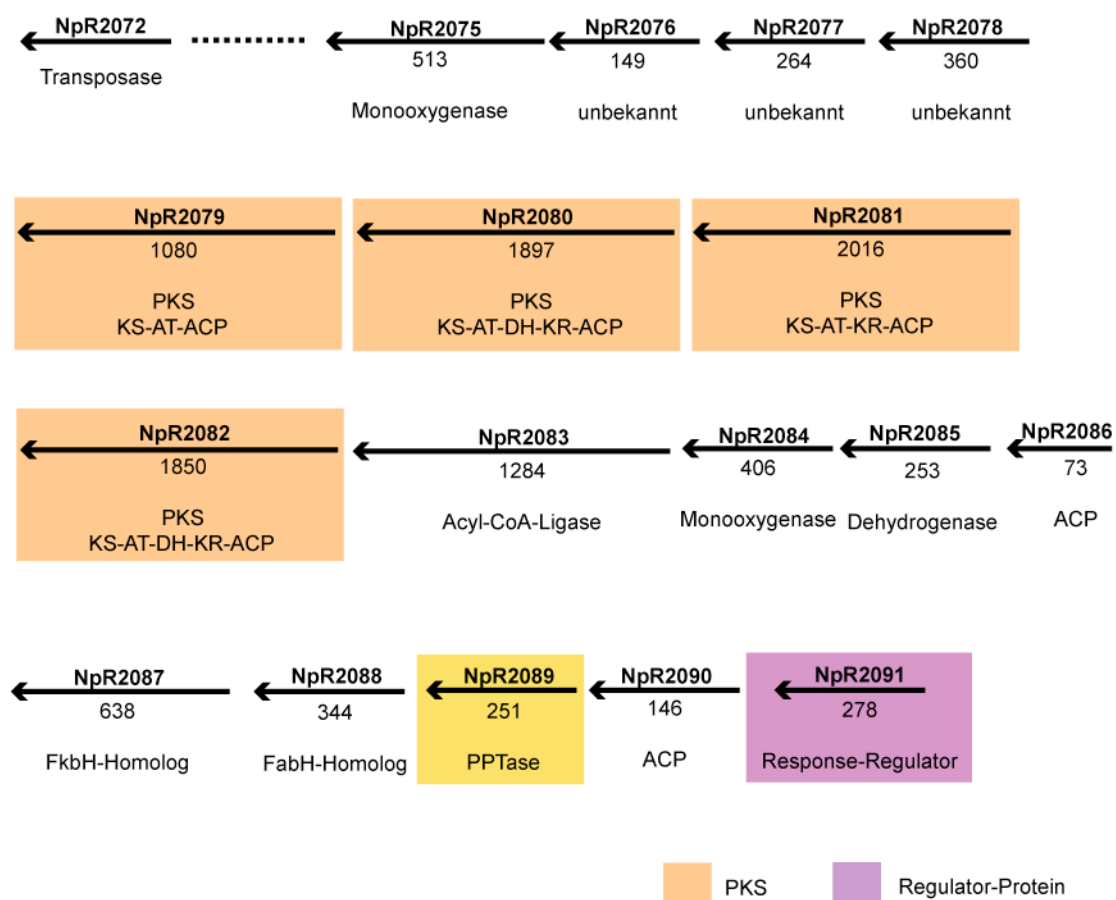


Abb. 30: Cluster III.

Die Pfeilrichtungen beziehen sich auf die Orientierung der ORFs im Genom. Über den Pfeilen ist die Gennummer angegeben, direkt unter den Pfeilen die Länge der codierten Proteine in Aminosäure-Resten. Danach folgt die aus Sequenz-Homologien abgeleitete Annotation. Im Fall der PKS ist darüber hinaus auch die Domänen-Zusammensetzung verzeichnet. PKS und Transporter-Proteine sind farblich hervorgehoben. Eng mit dem Cluster assoziiert liegt ein ORF, der für eine PPTase codiert (gelb hervorgehoben).

Cluster IV (**Abb. 31**) codiert für eine Hybrid-Synthetase aus NRPS und PKS. Es konnte aufgrund der Sequenz-Ähnlichkeiten zum entsprechenden Gencluster aus *Nostoc* sp. GSV224 als Nostopeptolid-Cluster identifiziert werden. Der Aufbau ist völlig identisch sowohl im Hinblick auf die Zahl der beteiligten Proteine als auch hinsichtlich deren Domänen-Struktur und Substratspezifitäten (vgl. (Hoffmann *et al.* 2003)). Übereinstimmend sind auch die Gene *nosE* und *nosF*, deren Genprodukte der Biosynthese der modifizierten Aminosäure 4-Methylprolin dienen (Luesch *et al.* 2003), und der durch *nosG* codierte ABC-Transporter. Das unbekannte Protein NpF2185 wird im Cluster von *Nostoc* sp. GSV224 ebenfalls codiert. Die Längen der Proteine NpF2185, NpF2186/NosE, NpF2187/NosF und NpF2188/NosG stimmen exakt überein, während sich bei den größeren NRPS- und PKS-Proteinen NpF2181/NosA, NpF2182/NosB und NpF2184/NosD Längenunterschiede zwischen 11 und 85 Aminosäurenresten zeigten. Die einzige markante Abweichung zwischen den beiden Nostopeptolid-Clustern besteht darin, dass NpF2183/NosC aus *N. punctiforme* im zweiten NRPS-Modul eine E-Domäne besitzt, die in NosC aus *Nostoc* sp. GSV224 nicht vorkommt. Direkt stromaufwärts von NpF2181 ist ein Transposase-Gen lokalisiert. In einem Abstand von ca. 12,7 kbp liegt stromaufwärts ein weiterer für eine PPTase codierender ORF (NpF2174).

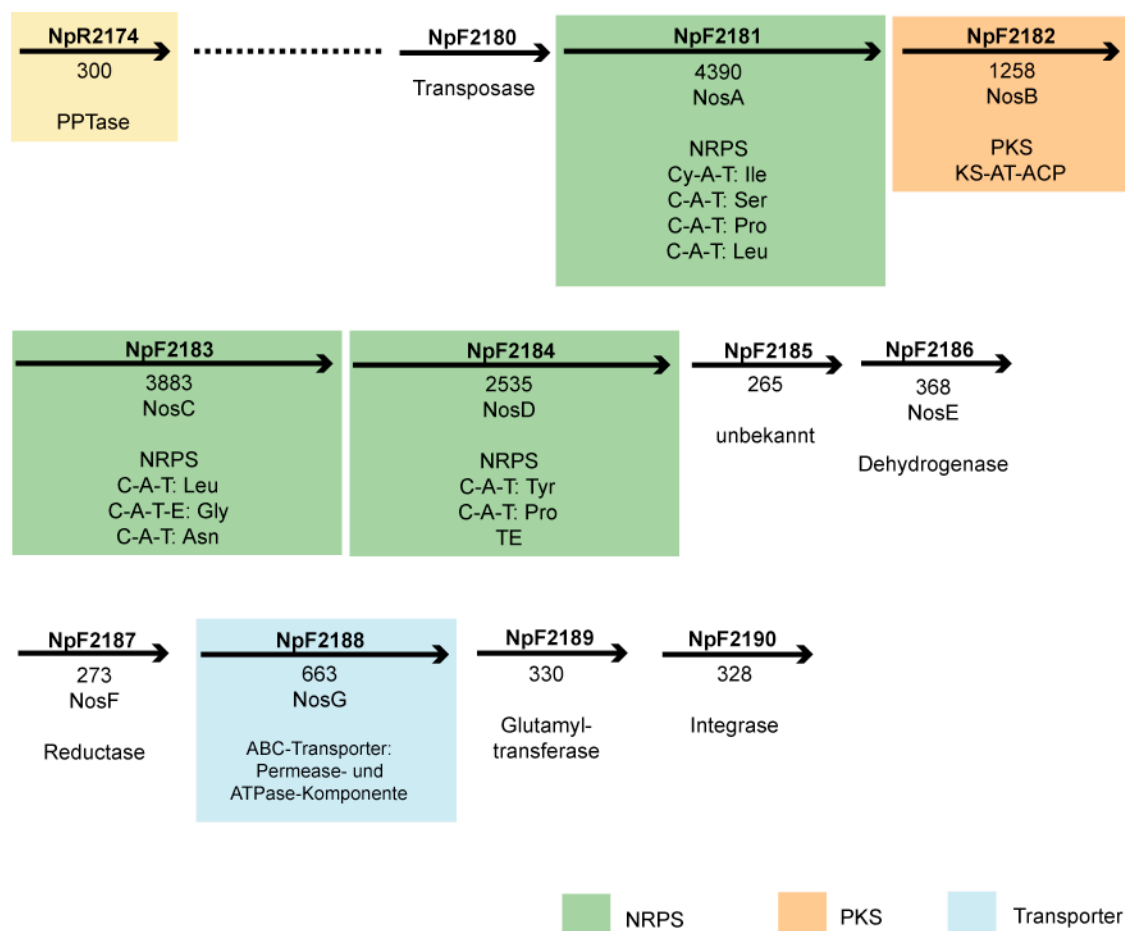


Abb. 31: Cluster IV: Nostopeptolid-Cluster.

Legende zu Abb. 31:

Die Pfeilrichtungen beziehen sich auf die Orientierung der ORFs im Genom. Über den Pfeilen ist die Gennummer angegeben, direkt unter den Pfeilen die Länge der codierten Proteine in Aminosäure-Fortsetzung der Resten. Danach folgt die aus Sequenz-Homologien abgeleitete Annotation. Im Fall der PKS und NRPS ist darüber hinaus auch die Domänen-Zusammensetzung verzeichnet. PKS, NRPS und Transporter-Proteine sind farblich hervorgehoben. Stromaufwärts in einem Abstand von etwa 12,7 kbp befindet sich ein ORF für eine PPTase (gelb hervorgehoben).

Cluster V (**Abb. 32**) stellt das einzige reine NRPS-Cluster in *N. punctiforme* dar. Die codierten NRPS-Enzyme zeigen einen regelgerechten Aufbau, beginnend mit NpF2460, das als erstes Modul ein typisches Beladungsmodul ohne C-Domäne aufweist, und abschließend mit NpF2463, das C-terminal eine TE-Domäne trägt. Das Cluster codiert ebenso wie das Nostopeptolid-Cluster für ein ABC-Transporter-Protein, das sowohl die Permease- als auch die ATPase-Komponente enthält.

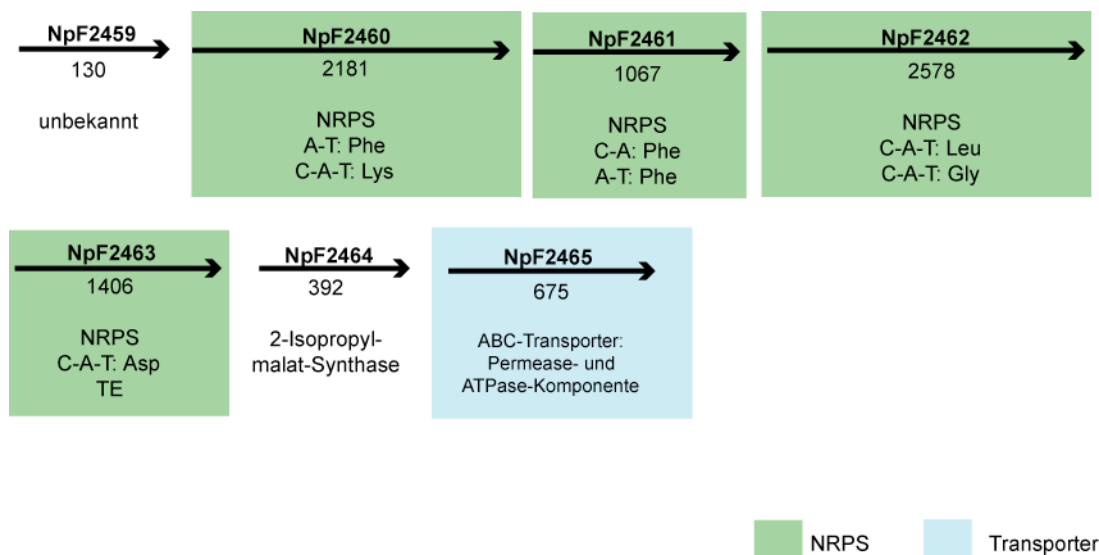


Abb. 32: Cluster V.

Die Pfeilrichtungen beziehen sich auf die Orientierung der ORFs im Genom. Über den Pfeilen ist die Gennummer angegeben, direkt unter den Pfeilen die Länge der codierten Proteine in Aminosäure-Resten. Danach folgt die aus Sequenz-Homologien abgeleitete Annotation. Im Fall der NRPS ist darüber hinaus auch die Domänen-Zusammensetzung verzeichnet. NRPS und Transporter-Proteine sind farblich hervorgehoben.

Das Cluster VI (**Abb. 33**) ist ein gemischtes Cluster mit NRPS- und PKS-Komponenten. Neben einem einmodularen PKS-Protein (NpR3024) und drei NRPS-Proteinen (NpR3023, NpR3025 und NpR3026) mit drei bis vier Modulen codiert es auch für ein Hybrid-Protein (NpR3027). Dieses ist mit einer Länge von 5352 Aminosäure-Resten, die einer berechneten molekularen Masse von 595 kDa entsprechen, das größte Protein aus der Klasse der PKS und NRPS in *N. punctiforme*. Auf der Polypeptidkette liegt das vermutliche NRPS-Beladungsmodul mit dem Domänen-Aufbau A-T, worauf ein PKS-Modul der Basiskonfiguration KS-AT-ACP folgt.

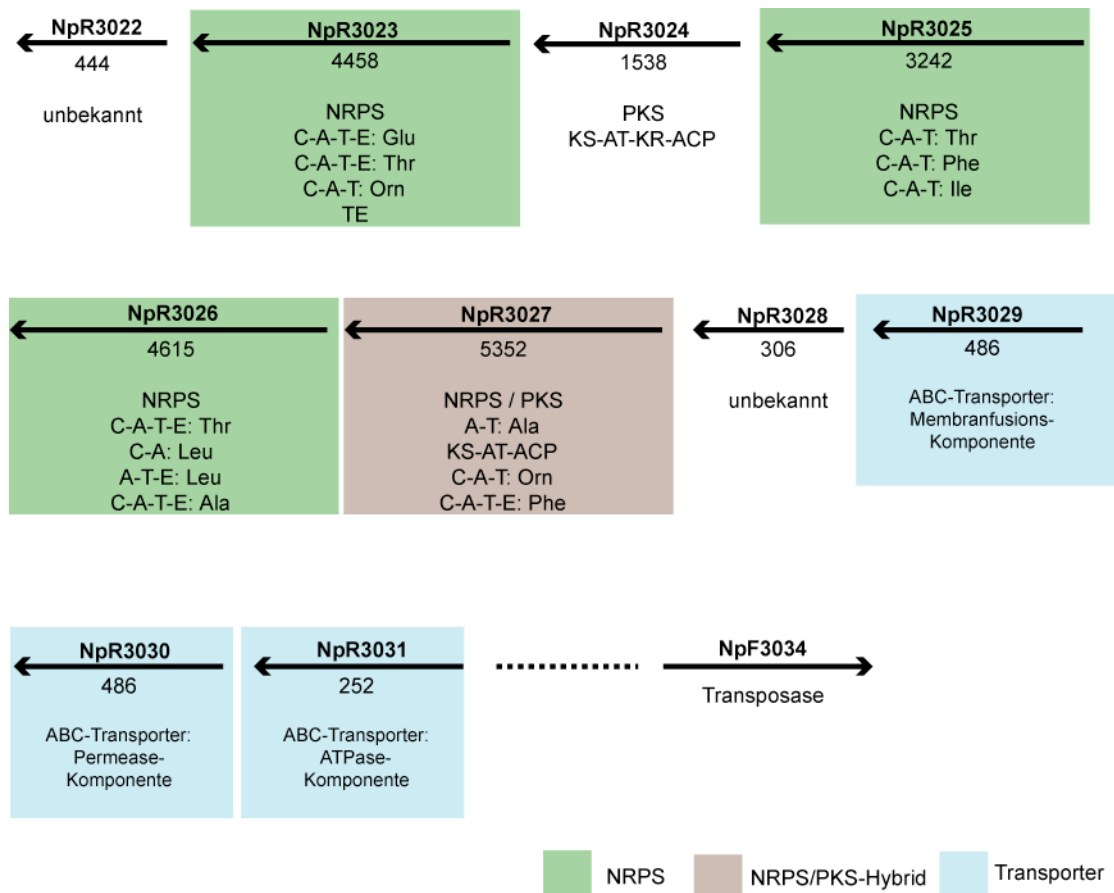


Abb. 33: Cluster VI.

Die Pfeilrichtungen beziehen sich auf die Orientierung der ORFs im Genom. Über den Pfeilen ist die Gennummer angegeben, direkt unter den Pfeilen die Länge der codierten Proteine in Aminosäure-Resten. Danach folgt die aus Sequenz-Homologien abgeleitete Annotation. Im Fall der PKS und NRPS ist darüber hinaus auch die Domänen-Zusammensetzung verzeichnet. PKS, NRPS, NRPS/PKS-Hybride und Transporter-Proteine sind farblich hervorgehoben.

Das Cluster VII (**Abb. 34**) gehört mit einer Länge von fast 63 kbp zu den besonders großen Clustern dieser Art in *N. punctiforme*. Es umfasst sieben PKS-Proteine mit insgesamt neun Modulen und ein einmodulares NRPS-Protein.

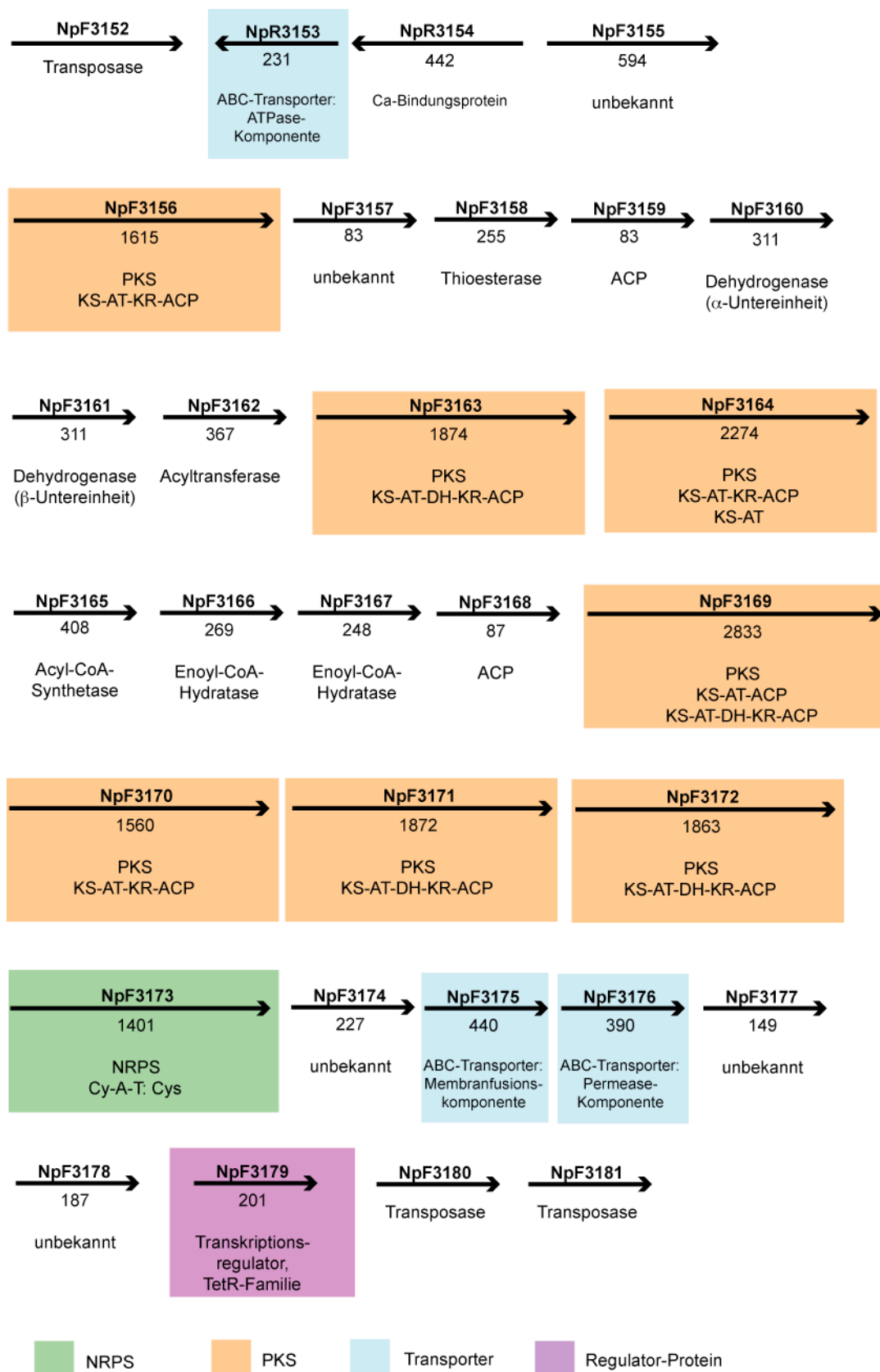


Abb. 34: Cluster VII.

Die weiteren Cluster (VIII, IX, X und XI) sind in **Abb. 35**, **Abb. 36**, **Abb. 37** und **Abb. 38** dargestellt.

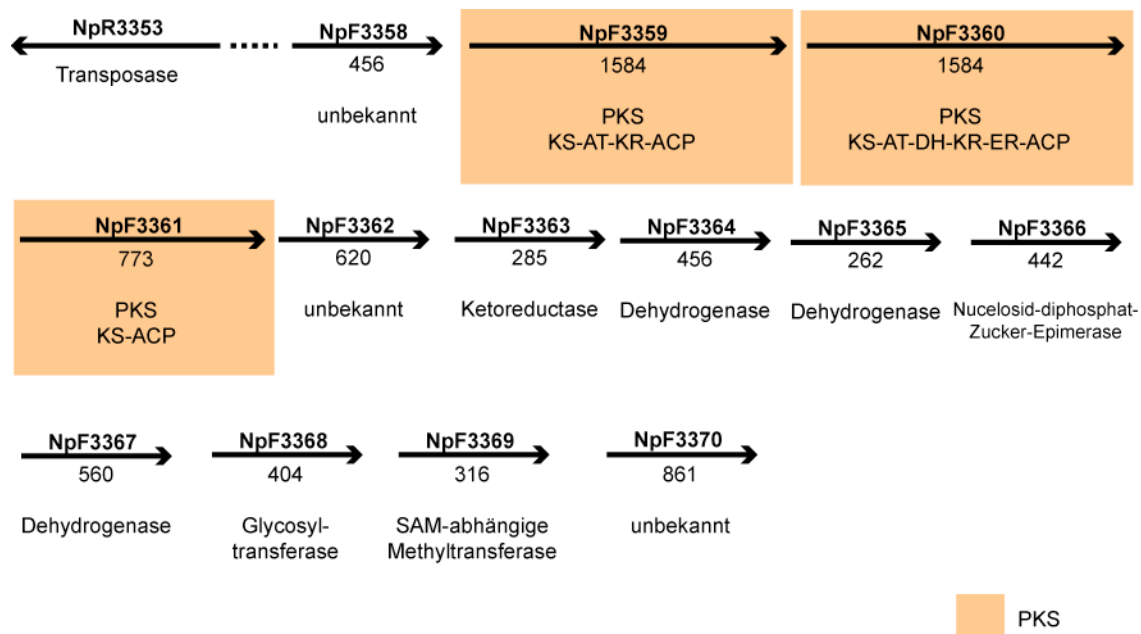


Abb. 35: Cluster VIII

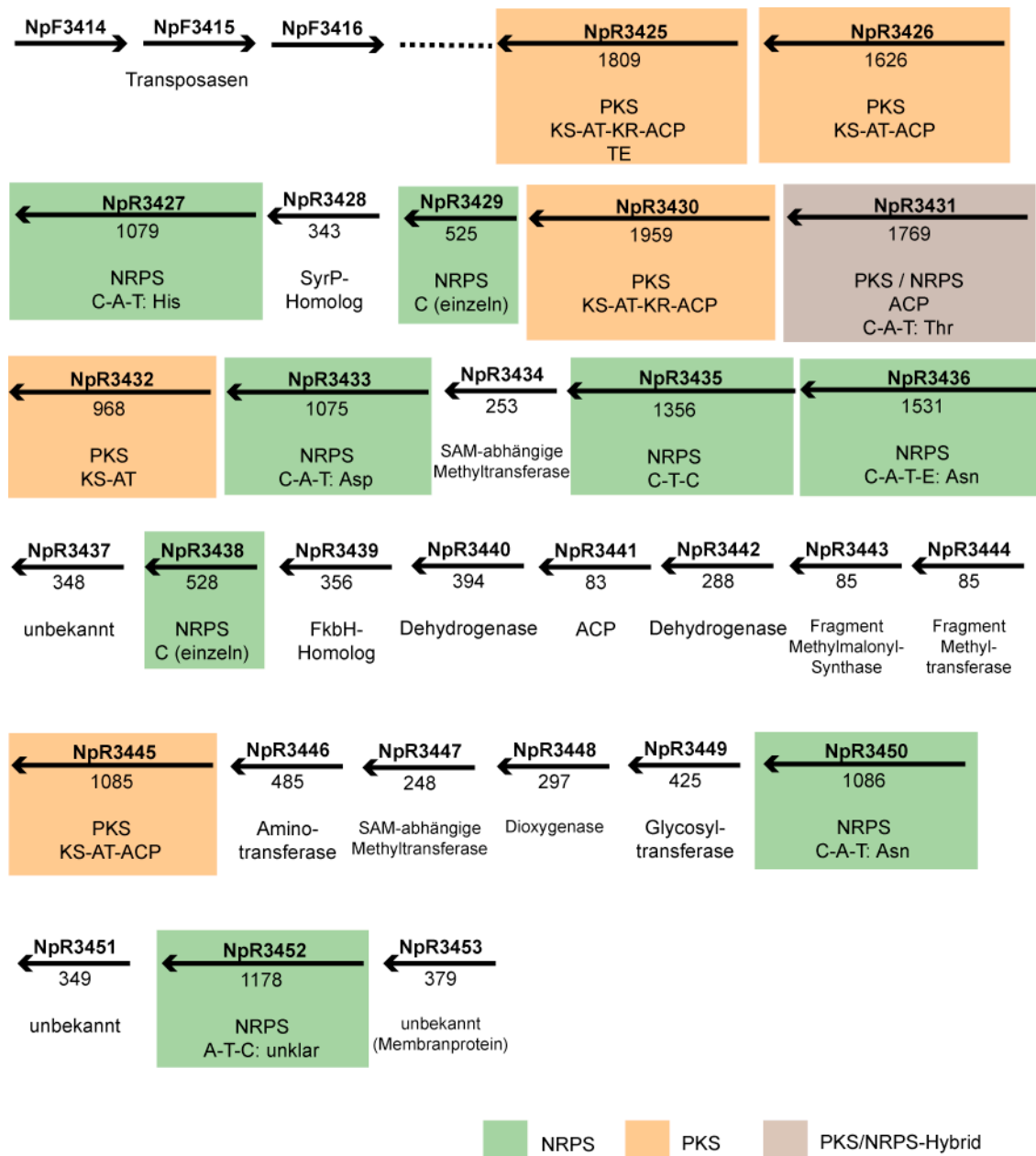


Abb. 36: Cluster IX.

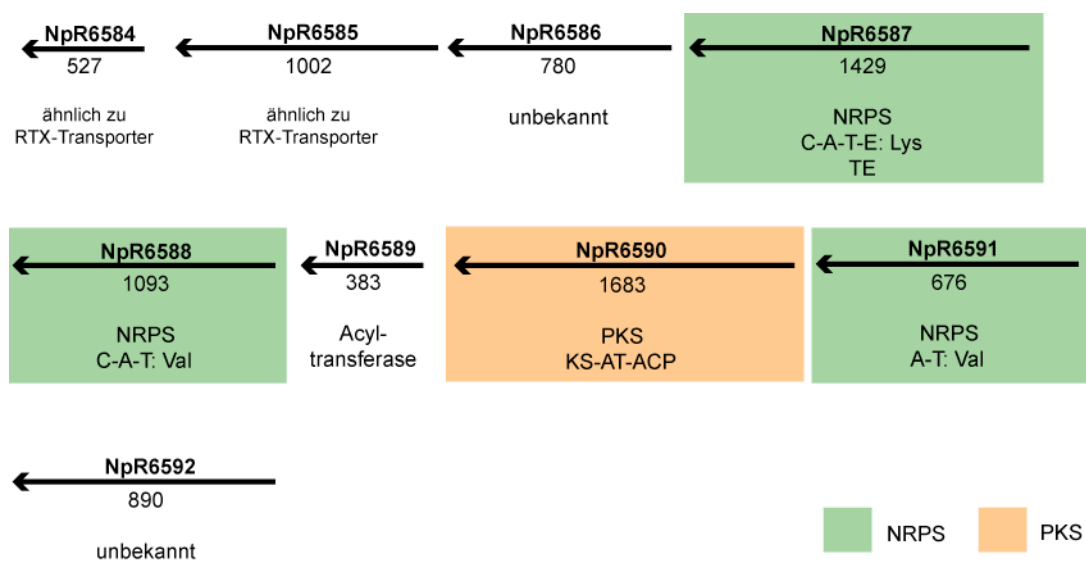


Abb. 37: Cluster X

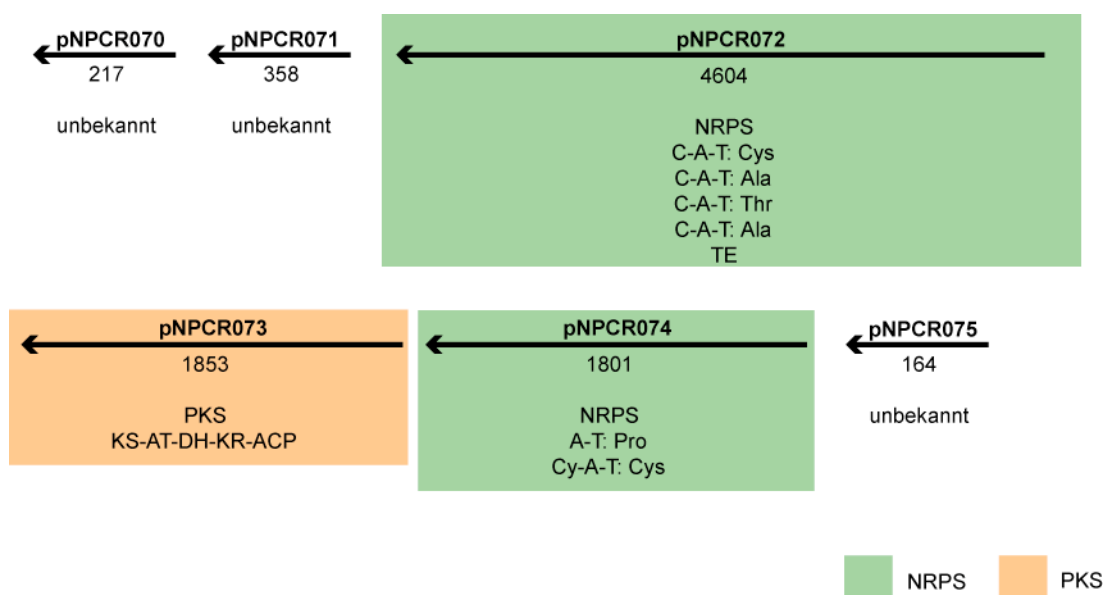


Abb. 38: Cluster XI.

3.5 Analytischer Nachweis von Polyketid- und Peptid-Sekundärmetaboliten in *Nostoc punctiforme*

Da in der Genom-Sequenz von *N. punctiforme* eine Reihe von PKS- und NRPS-codierenden Clustern identifiziert werden konnten, wurden die Zellen und der Kulturüberstand dieses Cyanobakteriums auf das Vorhandensein von Sekundärmetaboliten der entsprechenden Stoffgruppen untersucht. Dazu wurden sowohl Flüssigkulturen als auch auf Agarplatten gewachsenes Zellmaterial verwendet, wobei als Anzuchtmedium grundsätzlich das so genannte Allen & Arnon-Medium (AA-Medium) zum Einsatz kam (Allen und Arnon 1955). Bei den Anzuchtbedingungen wurde unterschieden zwischen Medium, das mit Ammoniumchlorid als Stickstoff-Quelle versehen war, und Medium ohne Zusatz von Ammoniumchlorid, bei dem rasch eine Stickstoff-Limitierung unter Ausbildung von Heterocysten eintritt. Außerdem wurden unterschiedliche Aufschluss- und Konzentrierungsverfahren eingesetzt und miteinander verglichen.

Bezüglich der unterschiedlichen Extraktionsverfahren für das Zellmaterial war festzustellen, dass die besten Ergebnisse mit einem Aufschluss der Zellen durch Ultraschallbehandlung in wässrigem Medium und nachfolgender Festphasenextraktion mittels C18-Kartuschen erzielt werden konnten. Im Vergleich zum häufig praktizierten Aufschluss in Methanol oder Methanol-Wasser-Gemischen und anschließendem Einengen der Lösungen wurden im HPLC-Chromatogramm dieselben Peakmuster gefunden, die Stärke der Signale war bei den wässrig aufgeschlossenen Proben jedoch deutlich größer (Daten nicht gezeigt). Zwischen den Proben aus Medium mit oder ohne Zusatz von gebundenem Stickstoff ließen sich in analytischen HPLC-Läufen keine klaren Unterschiede im Peakmuster feststellen, d. h. insbesondere traten keine zusätzlichen Signale oder deutliche Veränderungen in den relativen Signalstärken auf. Allerdings war auffällig, dass Zell- und Mediumüberstands-Proben von Stickstoff-limitierten Kulturen größere Mengen der Metabolite enthielten.

Für eine Übersichtsanalyse wurden wässrige Extrakte von *N. punctiforme*-Zellen, die in Schüttelkultur angezogen worden waren, und außerdem die nach der Zellernte angefallenen Zentrifugationsüberstände mittels Festphasenextraktion um den Faktor 1000 aufkonzentriert. Diese Proben wurden ohne weitere Extraktions- oder Reinigungsschritte in analytischen HPLC-Läufen auf einer „SymmetryShield“-C18-Säule (Waters) aufgetrennt. Während des Chromatographielaufs wurden in den relevanten Bereichen Fraktionen gesammelt, die in Zusammenarbeit mit der Technischen Universität Berlin (Dr. Martin Welker und Dr. Hans von Döhren) massenspektroskopisch mit dem MALDI-TOF-Verfahren untersucht wurden.

Im Zellpellet konnten die Masse-Ladungs-Quotienten m/z von sieben Verbindungen ermittelt werden, wobei von einigen auch *Post source decay*-Spektren (PSD-Spektren) erhalten wurden, deren Fragmente weitere Auskunft über die zugrunde liegenden Strukturen gaben. Im Mediumüberstand ließen sich die Masse-Ladungs-Quotienten von zwei Substanzen bestimmen. Die Daten und näheren Details hierzu sind in **Tab. 12** zusammengefasst.

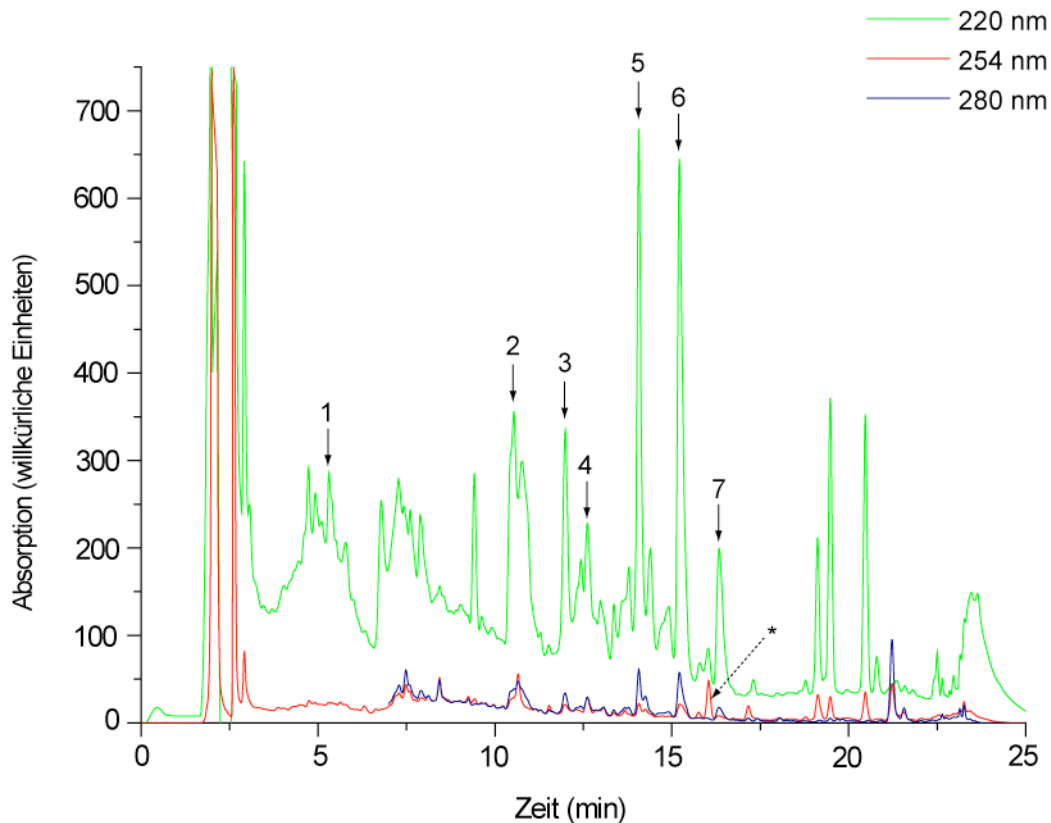


Abb. 39: „Reversed Phase“-HPLC-Auftrennung des Zellextrakts von *N. punctiforme*.

Gezeigt ist das Chromatogramm der registrierten Absorptionssignale bei Wellenlängen von 220 nm (grün), 254 nm (rot) und 280 nm (blau). Die Auftrennung erfolgte mit einem linearen Wasser-Acetonitril-Gradienten. Peaks, für die mittels MALDI-TOF-Massenspektrometrie eine Masse bestimmt werden konnte, sind mit Zahlen bezeichnet und in **Tab. 12** näher erläutert. Das Sternchen markiert den Peak, der im Chromatogramm des Überstandes besonders auffiel (vgl. **Abb. 40**).

Der Peak Nr. 3 ließ sich aufgrund der gefundenen Masse von 1081 m/z und den Fragmenten des PSD-Spektrums eindeutig einer der Nostopeptolid-Varianten A1 oder A3 zuordnen, die in *N. punctiforme* bereits identifiziert wurde (Hunsucker *et al.* 2004). Eine genauere Zuordnung zu einer der beiden Varianten, welche die gleiche Masse besitzen und sich nur in der Stellung einer Isobutyl-Gruppe unterscheiden, war nicht möglich. Weiterhin konnten für die Peaks Nr. 2, 5 und 7 aus dem Bakterienzellextrakt im PSD-Spektrum jeweils übereinstimmende Fragmente mit der dem Peak Nr. 3 zugrunde liegenden Substanz festgestellt werden, so dass es sich hierbei wahrscheinlich um diverse Nostopeptolid-Derivate mit leichten Modifikationen handelt. Dabei tritt das Nostopeptolid-Derivat des Peaks Nr. 2 mit einer Masse von 1075 m/z als einziges in deutlich detektierbaren Mengen auch im Kulturüberstand auf (Peak Nr. 9 in **Abb. 40**).

Tab. 12: Im Zellpellet und Mediumüberstand von *N. punctiforme*-Kulturen identifizierte Masse-Ladungs-Quotienten (vgl. Chromatogramme in Abb. 39 und Abb. 40)

Peak-Nr.	Quelle	m/z von [M+H] ⁺	Anmerkungen
1	Zellpellet	1042	Unbekanntes Peptid mit Valin, Leucin und Tryptophan
2	Zellpellet	1075	Nostopeptolid-Derivat, gleiche Fragmente wie Nr. 5
3	Zellpellet	1081	Nostopeptolid A
4	Zellpellet	842	Unbekanntes Peptid mit Lysin, Phenylalanin und Methylphenylalanin
5	Zellpellet	1087	Nostopeptolid-Derivat, gleiche Fragmente wie Nr. 2
6	Zellpellet	675	---
7	Zellpellet	1115	Nostopeptolid-Derivat, gleiche Fragmente wie Nr. 5 und zusätzliche +28 Da-Fragmente
8	Medium	462	Keine Bildung von Immonium-Ionen, kein Peptid
9	Medium	1075	Nostopeptolid-Derivat, gleiche Fragmente wie Nr. 2

Im Extrakt des Zellpellets konnten für zwei der Peaks (Nr. 1 und Nr. 4 in **Abb. 39**) im PSD-Spektrum Fragmentmassen identifiziert werden, die auf Peptide schließen lassen. Danach enthält die Substanz aus Peak Nr. 4 die Aminosäuren Lysin, Phenylalanin und Methylphenylalanin. Das einzige Gencluster, das NRPS-Module mit passender Aminosäure-Spezifität der Adenylierungsdomänen enthält, ist das Cluster V (siehe **Abb. 32**). Die von diesem reinen NRPS-Cluster codierten Enzyme sollten gemäß der Sequenzuntersuchung der Module ein zyklisches Heptapeptid mit der Aminosäuren-Zusammensetzung Phenylalanin (dreimal), Lysin, Leucin und Aspartat synthetisieren. Die zu erwartende Molekülmasse dieses Peptids, unter Berücksichtigung einer siebenfachen Abspaltung von Wassermolekülen bei der Ausbildung der Peptidbindungen, beträgt ca. 856 Da. Somit beläuft sich die Massendifferenz auf 14 Da, die durch eine zusätzliche Methylierung zu erklären wäre, wie sie in Form eines Methylphenylalanin-Restes auch durch die gemessenen PSD-Fragmente bestätigt wird.

Die Substanz aus Peak Nr. 1 enthält die Aminosäuren Valin, Leucin und Tryptophan. Auffälligerweise kann diese Information mit keinem der im Genom von *N. punctiforme* identifizierten Clustern ohne weiteres in Einklang gebracht werden. Bedenkt man, dass die Vorhersagemöglichkeiten für eine Tryptophan-Spezifität immer noch schlecht sind und eine klare Unterscheidung zu den beiden anderen aromatischen Aminosäuren nicht gewährleisten (Challis, Ravel und Townsend 2000; Rausch *et al.* 2005), käme das Cluster VI als Urheber dieses Peaks in Betracht (vgl. **Abb. 33**). Allerdings wäre in diesem Fall aufgrund der hohen Zahl an codierten NRPS-Modulen die zu erwartende Molekülmasse mit etwa 1460 Da deutlich höher als der massenspektroskopisch bestimmte Wert von 1042 m/z .

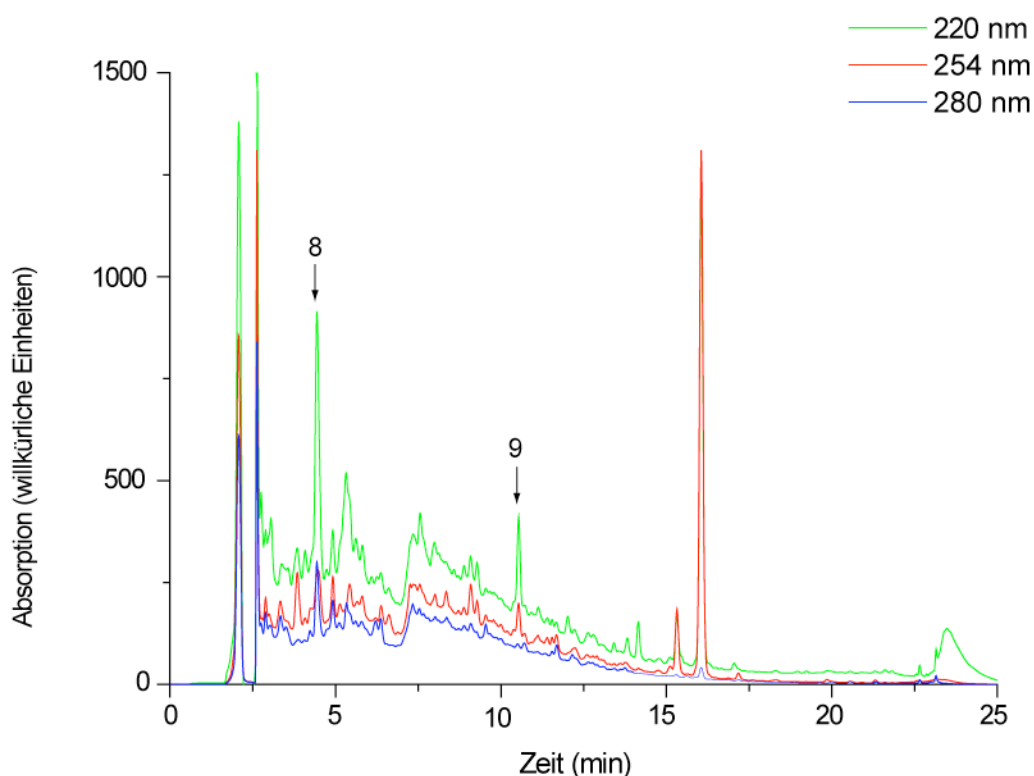


Abb. 40: „Reversed Phase“-HPLC-Auftrennung des Mediumüberstands von *N. punctiforme*

Gezeigt ist das Chromatogramm der registrierten Absorptionssignale bei Wellenlängen von 220 nm (grün), 254 nm (rot) und 280 nm (blau). Die Auftrennung erfolgte mit einem linearen Wasser-Acetonitril-Gradienten. Peaks, für die mittels MALDI-TOF-Massenspektrometrie eine Masse bestimmt werden konnte, sind mit Zahlen bezeichnet und in **Tab. 12** näher erläutert.

Im Kulturüberstand fand sich bei einer Laufzeit von etwa 16 min ein sehr prominenter Peak mit einer charakteristisch hohen Absorption bei 254 nm, der immer mit einem kleineren etwas früher eluierenden Begleitpeak auftrat (**Abb. 40**). Ein entsprechendes, jedoch weitaus schwächeres Absorptionssignal war auch im Chromatogramm des Zellextraktes zu beobachten (**Abb. 39**, mit einem Sternchen markiert). Von dieser Substanz, die offensichtlich überwiegend ins Kulturmedium abgegeben wurde und zu einem nur kleinen Anteil innerhalb der Zellen oder an der Zelloberfläche verblieb, konnten mittels MALDI-TOF-Massenspektroskopie keine Daten zum Masse-Ladungs-Verhältnis gewonnen werden.

3.6 Experimentelle Charakterisierung eines PKS/NRPS-Clusters aus *Nostoc punctiforme*

Für die detaillierte experimentelle Untersuchung eines PKS/NRPS-Biosynthese-Clusters wurde das Cluster VII ausgewählt, da es (1) zu den größten Clustern dieser Gruppe in *N. punctiforme* gehört, (2) ein Misch-Cluster darstellt, das sowohl für PKS als auch NRPS codiert, und (3) darüber hinaus die Sequenz eines Transkriptionsregulators enthält (siehe **Abb. 34**). Untersucht wurden die mögliche Beteiligung des Regulator-Proteins NpF3179 an der Steuerung der Expression von Komponenten des Clusters und die

enzymatische Aktivität und Spezifität der NRPS-Komponente NpF3173. Weiterhin wurden *Knockout*-Mutanten hergestellt, um den oder die Metaboliten, die von den Enzymen dieses Clusters produziert werden, im analytischen Vergleich von Extrakten des Wildtyps und der Mutanten identifizieren zu können.

3.6.1 Charakterisierung des Transkriptionsregulators

Zur Charakterisierung des Transkriptionsregulators NpF3179 wurde das Protein zunächst in den Zusammenhang bakterieller Transkriptionsregulatoren eingeordnet und seine Sequenz mit anderen Vertretern derselben Regulatorgruppe verglichen. Im nächsten Schritt wurde untersucht, ob innerhalb des Genclusters VII mögliche DNA-Bindungsstellen liegen. Da eine solche Bindungsstelle identifiziert werden konnte, wurde das Regulator-Protein heterolog in *Escherichia coli* expremiert, um die Bindung an den entsprechenden DNA-Abschnitt experimentell mittels eines *Electrophoretic mobility shift assay* (EMSA) zu verifizieren.

3.6.1.1 Sequenz-Untersuchung und phylogenetische Einordnung des Transkriptionsregulators

Aufgrund der Sequenz-Homologie ließ sich das Protein NpF3179 eindeutig der TetR-Familie von Transkriptionsregulatoren zuordnen. Sein Molekulargewicht berechnet sich aus der bekannten Aminosäure-Sequenz zu 18,7 kDa. Vertreter dieser Regulator-Familie finden sich in zahlreichen Bakteriengruppen, wobei sie besonders häufig in Actinobakterien, Bacillus-Stämmen und Pseudomonaden vorkommen (Ramos *et al.* 2005). Regulatoren vom TetR-Typ binden als Homodimere an die regulatorischen Sequenzen der Ziel-DNA (Orth *et al.* 2000). **Abb. 41** zeigt das Aminosäure-Sequenz-Alignment des Regulators aus *N. punctiforme* mit dem TetR-Regulator aus *Salmonella typhimurum*. Die mit Hilfe des Sekundärstruktur-Vorhersage-Programms PSIPRED ermittelten α -Helices in NpF3179 stimmen in ihrer Länge und Position über weite Teile des Proteins sehr gut mit den aus der Kristallstruktur bekannten Helices von TetR (Hinrichs *et al.* 1994) überein. Die N-terminalen Helices H2, H3 und H4 bilden die DNA-bindende Domäne, während die Helices H5 bis H10 Teile der regulatorischen Domäne sind (Orth *et al.* 2000). Die deutliche Abweichung im Bereich der Helix H7 korreliert mit dem Umstand, dass diese Region der zentrale Teil der Tetracyclin-Bindungstasche ist (Orth *et al.* 2000). Daneben ist NpF3179 im Vergleich zu TetR aus *S. typhimurum* etwa 30 Aminosäure-Reste kürzer, insbesondere fehlen die ersten ca. 25 Aminosäure-Reste des N-Terminus, von denen die Helix H1 ausgebildet wird.

Trotz der eindeutigen Zuordnung zu den TetR-Regulatoren erwies sich die Sequenz von NpF3179 in mehrfacher Hinsicht als ungewöhnlich. Blast-Suchen in der Genom-Datenbank von *N. punctiforme* ergaben, dass im Genom vier weitere Transkriptionsregulatoren vom TetR-Typ codiert werden (NpF3704, NpF6380, NpR1931 und NpR3597). Diese weisen im Gegensatz zu NpF3179 nicht die beschriebene N-terminale Verkürzung auf. NpF3179 unterscheidet sich bezüglich der Aminosäure-Zusammensetzung deutlich von den vier anderen TetR-Regulatoren: während diese untereinander eine Sequenz-Ähnlichkeit im Bereich von 50 % bis 65 % haben, wurden für NpF3179 Werte von nur 36 % bis 38 % gefunden. Eine allgemeine BlastP-Suche über die NCBI-Plattform mit der Sequenz von NpF3179 als Suchsequenz

erbrachte als beste Treffer eine Reihe von Proteinen aus Proteo- und Actinobakterien, allerdings keine aus Cyanobakterien (Tab. 13).

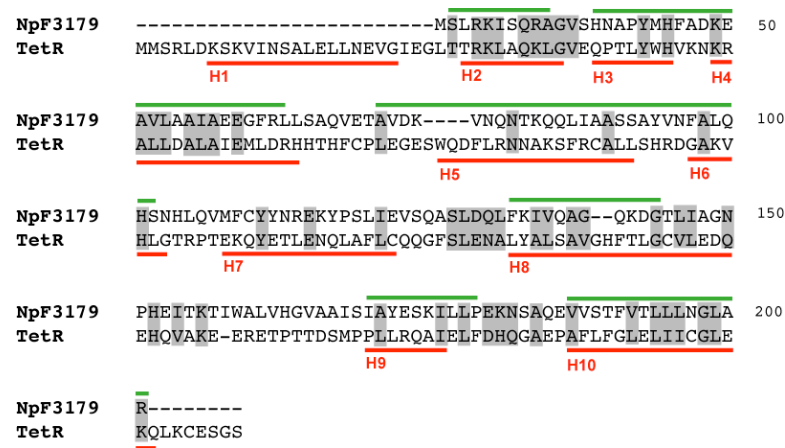


Abb. 41: Aminosäure-Sequenz-Alignment von TetR aus *Salmonella typhimurum* und dem putativen Transkriptionsregulator NpF3179 aus *N. punctiforme*.

Grau unterlegt sind Positionen, die in beiden Sequenzen identische oder ähnliche Aminosäure-Reste aufweisen. In der Sequenz aus *Salmonella* sind mit roten Strichen die aus der Kristallstruktur bekannten Helix-Bereiche H1 bis H10 markiert und entsprechend benannt. Die grünen Striche geben die für das Protein NpF3179 mittels PSIPRED vorhergesagten Helix-Bereiche an.

Tab. 13: Beste Treffer der BlastP-Suche mit NpF3179 als Suchsequenz

Organismus	Bakteriengruppe	Accession-Nummer
<i>γ</i> -Proteobakterium KT 71	<i>γ</i> -Proteobakterien	ZP_01101208
<i>Oceanobacter</i> sp. RED65	<i>γ</i> -Proteobakterien	ZP_01308386
<i>Streptomyces avermitilis</i> MA-4680	Actinobakterien	NP_826261
<i>Bdellovibrio bacteriovorus</i> HD100	<i>δ</i> -Proteobakterien	CAE81084
<i>Marinomonas</i> sp. MED121	<i>γ</i> -Proteobakterien	ZP_01074415
<i>Hahella chejuensis</i>	<i>γ</i> -Proteobakterien	YP_436875
<i>γ</i> -Proteobakterium (nicht kultiviert)	<i>γ</i> -Proteobakterien	AAS73066
<i>Gluconobacter oxydans</i> 621H	<i>α</i> -Proteobakterien	AAW62109
<i>Mycobacterium vanbaalenii</i> PYR-1	Actinobakterien	ZP_01207510
<i>Rhodobacterales bacterium</i> HTCC2654	<i>α</i> -Proteobakterien	ZP_01015676
<i>Nocardia farcinica</i> IFM 10152	Actinobakterien	BAD57532

Eine phylogenetische Analyse dieser Sequenzen, ergänzt um weitere TetR-Regulatorproteine aus *N. punctiforme* und anderen Bakterien, ergab den Baum in Abb. 42. Daraus ist erkennbar, dass NpF3179 nicht mit den cyanobakteriellen Transkriptionsregulatoren der TetR-Familie zusammenfällt, sondern in einem deutlich abgegrenzten Cladus liegt, der im übrigen aus Sequenzen von *γ*-Proteobakterien und einem *δ*-Proteobakterium besteht. Wenngleich die *Clade probability values* in diesem Bereich des Baums durchgehend niedrig ausfielen, konnten alle basalen Aufspaltungen mit dem NJ-Verfahren und MP-

vorgegebenen Suchsequenz konnte BPROM einen potenziellen Bindungsort für einen TetR-Transkriptionsfaktor identifizieren. Dieser Bindungsort erhielt einen hohen Scoring-Wert, d. h. er wurde hinsichtlich der Sequenz und im Zusammenhang mit den in der Nähe liegenden Promotorsequenzen als sehr wahrscheinlich richtig eingestuft. Die Verhältnisse im identifizierten Sequenzbereich sind in **Abb. 43** dargestellt.

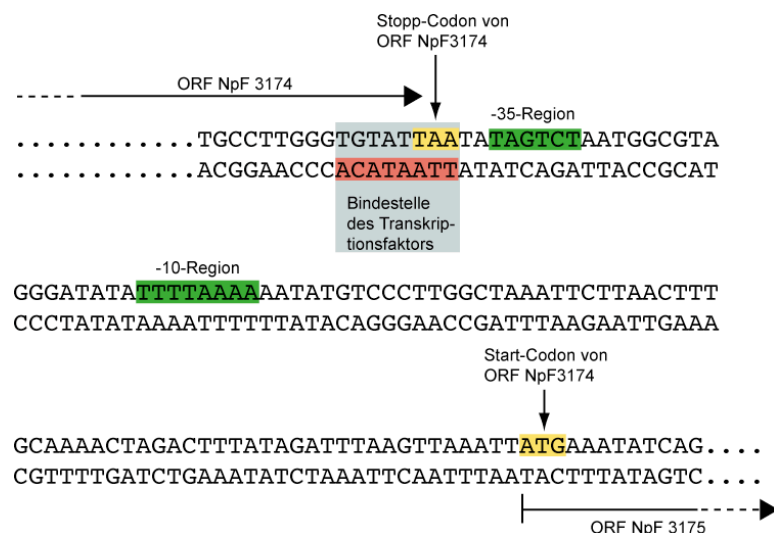


Abb. 43: Vorhergesagte Bindestelle des Transkriptionsfaktors vom TetR-Typ

Gemäß der Vorhersage des Programms BPROM der Firma Softberry befindet sich eine mögliche Bindestelle im Sequenzbereich zwischen den ORFs NpF3174 und NpF3175. Die Bindestelle TTAATACA (hier in 3'-5'-Richtung) ist rot unterlegt. Die Bindungsregionen des dazu gehörigen Promotors sind grün unterlegt. Stopp- und Startcodons sind gelb markiert.

Die identifizierten Sequenzelemente befinden sich im Bereich zwischen den ORFs NpF3174 und NpF3175. Letzterer codiert für die Membranfusionskomponente eines ABC-Transporters, wie in **Abb. 34** dargestellt. Die -10-Region des potenziellen Promotors liegt etwa 60 bp vor dem Start-Codon von ORF NpF3175, so dass sich für die nicht-translatierte Region der entsprechenden mRNA am 5'-Ende eine Länge von ca. 50 bp errechnet. Dieser Wert liegt im Rahmen der Abstände zwischen Transkriptionsstartpunkt und Startcodon, wie er für andere Cyanobakterien gefunden wurde (Vogel *et al.* 2003). Die mögliche Bindestelle des Transkriptionsfaktors ist direkt vor der -35-Region des Promotors gelegen und überschneidet sich mit dem Ende des ORF NpF3174. Bei der vorliegenden Konstellation ist davon auszugehen, dass der Regulator bei Besetzung seiner Bindestelle die Bindung der RNA-Polymerase an den Promotor von NpF3175 verhindert und somit in Bezug auf die Transkription der mRNA für die ABC-Transporterkomponente NpF3175 einen negativ regulatorischen Effekt ausübt. Dies steht in Übereinstimmung mit der Beobachtung, dass fast alle untersuchten Transkriptionsfaktoren der TetR-Familie als negative Regulatoren wirken (Ramos *et al.* 2005).

3.6.1.3 Heterologe Expression des Regulators NpF3179

Zur Vorbereitung der EMSA-Untersuchung wurde der Regulator NpF3179 in *Escherichia coli* TOP10-Zellen heterolog expremiert und aus dem Zellextrakt aufgereinigt. Hierzu wurde unter Verwendung der Primer „636/8-FWD“ und „636/8-REV“ die codierende Sequenz aus genomischer DNA von *N. punctiforme* mittels PCR amplifiziert, wobei eine Mischung aus Pfu-DNA-Polymerase mit *proofreading*-Aktivität und Taq-Polymerase zum Einsatz kam. Das PCR-Produkt wurde über eine TA-Klonierung in den Expressionsvektor pBAD-TOPO-TA (Invitrogen) kloniert. Bei der Expression mit Hilfe dieses Vektors wird dem Protein C-terminal ein 6x His-Tag angefügt. Die erhaltenen Vektorkonstrukte wurden durch vollständige Sequenzierung auf mögliche Basenaustausche oder andere Sequenzveränderungen untersucht. Für die heterologe Expression wurde ein Vektorkonstrukt gewählt, das in der Proteincodierenden Sequenz vollständig mit der Originalsequenz des Regulators NpF3179 übereinstimmte.

In einer Reihe von Testexpressionen wurde der Einfluss der Temperatur und der Konzentration des Induktors L-Arabinose auf die Menge an expremiertem Protein und auf dessen Lokalisierung in *inclusion bodies* untersucht. Dazu wurden die Zellysate der Testansätze zentrifugiert und die erhaltenen Überstände und Pellets getrennt mittels SDS-PAGE und Western Blots analysiert. Im Rahmen dieser Testexpressionen wurde ermittelt, dass eine Konzentration des Induktors L-Arabinose von 0,02 % (w/v) in Kombination mit einer Kultivierungstemperatur von 25° C besonders günstig war, da unter diesen Bedingungen bei starker Überexpression das Protein zum deutlich überwiegenden Anteil in der löslichen Zellfraktion lokalisiert war (Daten nicht gezeigt).

Nach Überexpression unter diesen Bedingungen wurde das C-terminal mit einem 6x His-Tag versehene Protein aus dem Zentrifugationsüberstand des Zellysats durch eine Affinitätschromatographie an Nickel-Agarose aufgereinigt, wobei die Elution des Proteins durch Steigerung der Imidazol-Konzentration erfolgte. Die Aufreinigung mit drei Waschschritten nach dem Standardprotokoll des Herstellers ergab weitgehende Reinheit der späteren Elutionsfraktionen gemäß der Beurteilung mit einer Coomassie-Färbung des Polyacrylamidgels (siehe **Abb. 44**).

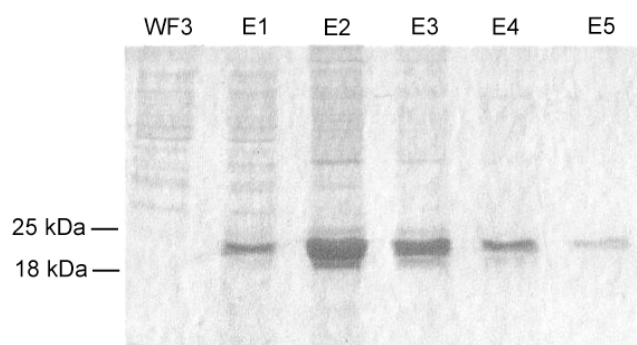


Abb. 44: Aufreinigung des His-Tag-markierten Regulatorproteins NpF3179 mittels Affinitätschromatographie an Ni-Agarose

Auf das SDS-PAGE-Gel wurden die dritte Waschfraktion (WF3) sowie Aliquots der verschiedenen Elutionsfraktionen (E1 bis E5) aufgetragen. Das Gel wurde nach dem Lauf mit Coomassie angefärbt.

Das Molekulargewicht des aufgereinigten Regulators betrug gemäß SDS-PAGE etwa 20 kDa, was in guter Übereinstimmung zum errechneten Wert von 19,7 kDa steht, der sich aus der Aminosäure-Sequenz des Proteins unter Berücksichtigung der sechs zusätzlichen Histidin-Reste ergibt. Bestimmungen der Proteinkonzentration mittels der Bradford-Methode ergaben, dass sich unter den angewandten Expressions- und Aufreinigungsbedingungen durchschnittlich gut reproduzierbar etwa 1 mg Protein pro 100 ml Kulturmedium gewinnen ließen. Nach Dialyse gegen 10 mM TBS konnte das Protein in Gegenwart von 30 % (w/v) Glycerol über mehrere Wochen bei –20° C gelagert werden, ohne dass sich in Coomassie-gefärbten Polyacrylamid-Gelen durch Abbau bedingte neue Banden feststellen ließen.

3.6.1.4 Untersuchung der DNA-Bindung des Regulators NpF3179 mittels EMSA

Um zu untersuchen, ob der Regulator NpF3179 *in vitro* an den vorhergesagten DNA-Abschnitt binden kann, wurde ein *Electrophoretic mobility shift assay* (EMSA) durchgeführt. Dieses Verfahren beruht darauf, das in Frage stehende Regulator-Protein gemeinsam mit DNA-Molekülen zu inkubieren, welche die mögliche Bindestelle enthalten, und anschließend den Ansatz in einem Polyacrylamid-Gel elektrophoretisch unter nativen Bedingungen aufzutrennen. Im Falle einer Bindung des Proteins an die DNA-Moleküle ist deren Wanderungsgeschwindigkeit im Gel im Vergleich zu freien DNA-Molekülen verzögert. Da der Regulator NpF3179 mit knapp 20 kDa ein relativ kleines Molekulargewicht aufweist, musste darauf geachtet werden, die Länge der verwendeten DNA-Fragmente angemessen zu gestalten. Die Länge wurde dementsprechend mit 138 bp relativ kurz bemessen, um Veränderungen der Wanderungsgeschwindigkeit mit ausreichender Sicherheit bestimmen zu können.

In der Untersuchung kamen zwei DNA-Fragmente zum Einsatz. Das eine, DNA (BIND), enthielt etwa mittig die vorhergesagte Bindestelle (siehe **Abb. 43**), das andere, als Negativprobe dienende Fragment, DNA (NONBIND), war zufällig aus der Sequenz des Cluster VII gewählt worden und enthielt diese Bindestelle nicht. Beide Fragmente wurden mittels PCR aus der genomischen DNA von *N. punctiforme* amplifiziert, aus dem PCR-Ansatz durch eine Agarose-Gelelektrophorese aufgereinigt und anschließend am 5'-Ende mit ³²P radioaktiv markiert.

Es konnte eine Bande identifiziert werden, die nur auftrat, wenn das Regulator-Protein NpF3179 und DNA (BIND), d. h. das DNA-Fragment mit Bindungsstelle, im Ansatz kombiniert wurden (siehe **Abb. 45**, Spuren 1 und 2). Diese Bande war hingegen nicht detektierbar, wenn statt dessen DNA (NONBIND) zugesetzt wurde (**Abb. 45**, Spur 8) oder der Ansatz ohne Regulator-Protein inkubiert worden war (**Abb. 45**, Spur 9). Zur weiteren Kontrolle wurden auch Nickel-Affinitäts-„Aufreinigungen“ aus Zelllysaten benutzt, die aus *E. coli* TOP10-Zellen mit leerem pBAD-TOPO-Vektor stammten und daher kein expremiertes Regulator-Protein enthalten konnten, um auszuschließen, dass die Bande durch ein anderes Protein des Lysats hervorgerufen wurde. Mit diesen Proben konnte keine Verschiebung der Bande unbesetzter DNA erzielt werden (Daten nicht gezeigt). Weiterhin konnte gezeigt werden, dass sich das Signal durch Zugabe von unmarkierter DNA (BIND) abschwächen ließ (**Abb. 45**, Spuren 3 bis 7), während diese Wirkung bei Verwendung von unspezifischem poly(dI-dC) erst mit bedeutend größeren Mengen eintrat.

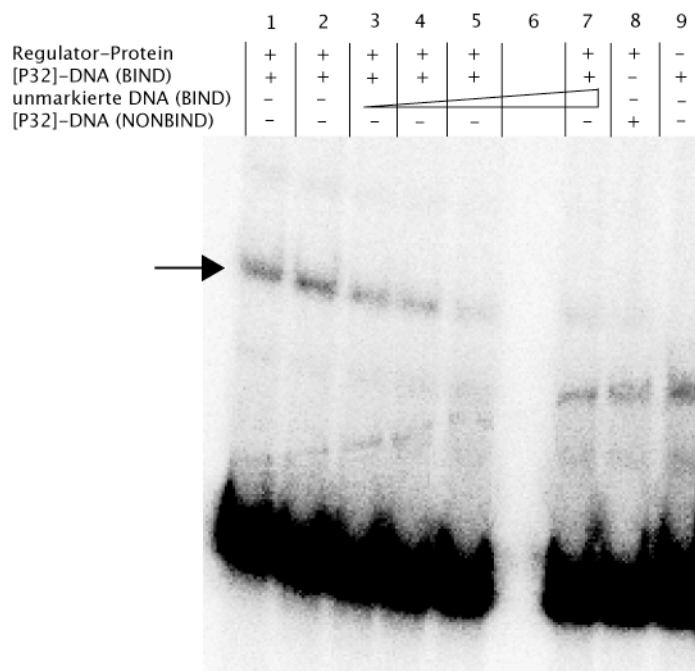


Abb. 45: Untersuchung der DNA-Bindung des Regulators NpF3179 mittels EMSA.

Die Ansätze wurden nach der Inkubation in einem nativen Polyacrylamid-Gel elektrophoretisch aufgetrennt. Anschließend wurden die radioaktiven Signale durch Exposition eines PhosphorImager-Schirms mit dem getrockneten Gel sichtbar gemacht. Die Zusammensetzung der Ansätze ist in dem Tabellenschema gezeigt. DNA (BIND) bezeichnet das DNA-Fragment mit Bindestelle, DNA (NONBIND) dasjenige ohne Bindestelle. Das Dreieck in der Tabelle gibt von links nach rechts ansteigende Mengen an. In die Spuren Nr. 1 und 2 wurden identische Proben aufgetragen, die Spur Nr. 6 wurde freigelassen. Der Pfeil markiert die durch Bindung des Regulator-Proteins hervorgerufene Bande.

Wie aus **Abb. 45**, die ein typisches Beispiel für die durchgeführten EMSA-Untersuchungen darstellt, ebenfalls erkennbar ist, konnte nur ein geringer Anteil der markierten DNA in der Laufgeschwindigkeit verzögert werden, während der ganz überwiegende Teil keinen „Shift“ zeigte. Diese Verschiebung war auch nur erkennbar, wenn ein erheblicher molarer Überschuss an Protein eingesetzt wurde, wie folgende Rechnung zeigt. Typischerweise wurden 50 ng Protein, also ca. $2,7 \times 10^{-12}$ mol Protein, im Inkubationsansatz mit etwa 12 ng DNA kombiniert, die bei einer Fragmentlänge von 138 bp und somit einem Molekulargewicht von 91 kDa etwa $1,3 \times 10^{-13}$ mol entsprachen. Allerdings konnte in unterschiedlichen Mengenkombinationen von Protein und DNA stets eine, wenn auch schwache „Shift“-Bande nachgewiesen werden, die spezifisch nur bei Zugabe von DNA (BIND) auftrat, so dass insgesamt deutliche Anhaltspunkte für eine spezifische Bindung des Transkriptionsregulators NPF3179 gewonnen werden konnten.

3.6.2 Charakterisierung des NRPS-Moduls NpF3173

Das Cluster VII aus *N. punctiforme* ist ein gemischtes PKS/NRPS-Cluster, das neben einer Reihe von PKS-Enzymen auch die Sequenz eines Proteins enthält, das zur Klasse der NRPS gehört (siehe **Abb. 34**). Der ORF NpF3173 codiert für ein komplettes NRPS-Modul, bestehend aus einer Kondensations-, einer

Adenylierungs- und einer Thiolierungsdomäne, wobei die Kondensationdomäne jedoch als Sonderform der Cyclisierungsdomäne (Cy-Domäne) ausgebildet ist (vgl. Abschnitt 1.3.2.2). Die Anwendung des Spezifitäts-Codes für Adenylierungsdomänen ergab, dass NpF3173 wahrscheinlich die Aminosäure Cystein aktiviert, was insofern zum Vorliegen einer Cy-Domäne in diesem Modul passt, als Cy-Domänen im Allgemeinen die Aminosäuren Serin, Threonin oder Cystein umsetzen. Um die Funktionalität dieses NRPS-Moduls und die aus der Sequenz abgeleiteten Vorhersagen zu überprüfen, wurde NpF3173 heterolog in *Escherichia coli*-Zellen expremiert, aus dem Zelllysate aufgereinigt und hinsichtlich seiner Substratspezifität biochemisch untersucht.

3.6.2.1 Heterologe Expression des NRPS-Moduls NpF3173 in *Escherichia coli*

Für die Überexpression in *E. coli* wurde die Sequenz des ORF NpF3173 zunächst mittels PCR aus genomischer DNA von *N. punctiforme* amplifiziert. Die bei der PCR eingesetzten Primer waren so konstruiert, dass das PCR-Produkt direkt mittels einer TA-Klonierung unter Einhaltung des Leserahmens in den pBAD-TOPO-Vektor (Invitrogen) kloniert werden konnte. Anschließend wurden zwei der erhaltenen Plasmide im Insert-Bereich vollständig über etwa 4200 Nucleotide sequenziert, um die Vollständigkeit und Richtigkeit der klonierten Sequenz zu gewährleisten. Das für die Expression benutzte Plasmid besaß ein Insert, dessen Sequenz exakt identisch mit der veröffentlichten Genom-Sequenz war. Dieses Expressionsplasmid wurde in *Escherichia coli* TOP10-Zellen eingeführt, in denen das Protein bei 25° C zur Überexpression gebracht wurde. Aufgrund des am C-Terminus des Proteins angefügten 6x His-Tags konnte es aus dem Zelllysate durch eine Affinitätschromatographie an Nickel-Agarose aufgereinigt werden. Nach diesem Reinigungsschritt war das Protein, beurteilt anhand von Coomassie-gefärbten SDS-Polyacrylamid-Gelen, bereits weitgehend homogen (**Abb. 46**). Das Molekulargewicht des aufgereinigten Proteins lag gemäß SDS-PAGE bei etwa 160 kDa und somit in guter Übereinstimmung mit dem aus der Sequenz errechneten Wert, der einschließlich der sechs zusätzlichen Histidin-Reste 157,6 kDa beträgt.

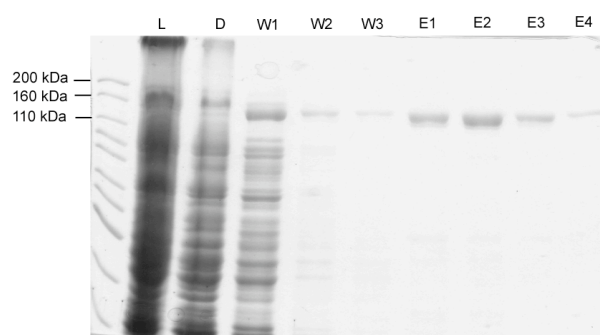


Abb. 46: Aufreinigung des NRPS-Moduls NpF3173 nach heterologer Expression in *E. coli*.

SDS-PAGE verschiedener Fraktionen der Affinitätsaufreinigung mit Nickel-Agarose. Aufgetragen wurden Aliquots des Lysats (L), des Säulendurchflusses (D), der drei Waschfraktionen (W1, W2 und W3) sowie des fraktionierten Eluates (E1, E2, E3 und E4). Färbung des Gels erfolgte mit Coomassie.

3.6.2.2 Adenylierungsspezifität des NRPS-Moduls NpF3173

Gemäßt der Vorhersage mit Hilfe des Spezifitäts-Codes für Adenylierungsdomänen (Stachelhaus, Mootz und Marahiel 1999; Challis, Ravel und Townsend 2000) setzt das NRPS-Modul NpF3173 aus dem Cluster VII die Aminosäure Cystein um, da die kritischen Aminosäurereste im aktiven Zentrum im Einbuchstaben-Code DLYNMSLIWK lauten. Diese aus der Sequenz abgeleitete Adenylierungsspezifität des NRPS-Moduls wurde mit dem ATP-Pyrophosphat-Austausch-Assay experimentell überprüft, der die reversible Bildung von Enzym-gebundenem Aminoacyl-AMP misst (Gevers, Kleinkauf und Lipmann 1969). Je mehr Aminoacyl-AMP gebildet wird, desto höher ist der Einbau von radioaktiv markiertem Pyrophosphat in ATP, was durch Szintillationszählung nachgewiesen werden kann. Für diese Untersuchungen wurden diejenigen Elutionsfraktionen aus frischen Protein-Expressionen und Aufreinigungen eingesetzt, in denen sich die größte Menge an Enzym befand. In mehreren unabhängig voneinander durchgeführten Assays wurde die Umsetzung von L-Cystein, L-Serin, L-Threonin und L-Alanin gemessen.

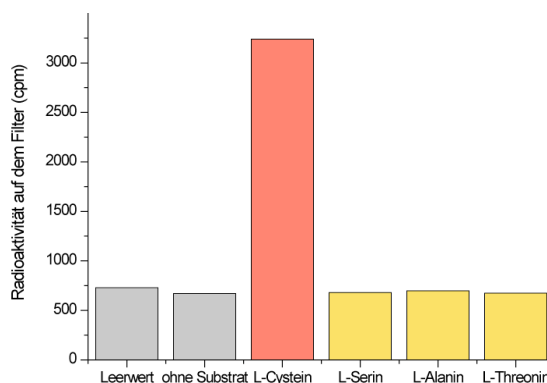


Abb. 47: Adenylierungsspezifität des NRPS-Moduls NpF3173

Gemessen wurde die durch den Einbau von [32 P]-Pyrophosphat in ATP verursachte Radioaktivität auf dem Filter. Der Leerwert wurde ohne Enzym und ohne Substratzugabe gemessen. Dargestellt sind die Durchschnittswerte aus Doppelansätzen eines Experiments.

Aus **Abb. 47** ist ersichtlich, dass die auf den Filtern verbliebene Radioaktivität für die Substrate L-Serin, L-Alanin und L-Threonin jeweils ähnliche Werte erreichte wie für den Leerwert, der ohne Enzym und ohne Substrat gemessen wurde, und wie für einen Ansatz ohne jegliches Aminosäure-Substrat. Mit diesen Substraten kam es also zu keiner Umsetzung. Hingegen konnten für L-Cystein durchschnittlich fünf- bis sechsfach höhere Werte ermittelt werden, weshalb davon auszugehen ist, dass NpF3173 entsprechend der vorhergesagten Aminosäurespezifität ausschließlich L-Cystein effektiv umzusetzen vermag.

3.6.3 Untersuchung einer Knockout-Mutation der NRPS-Komponente im Cluster VII

Das Gencluster VII codiert für eine kombinierte PKS/NRPS-Synthase (siehe **Abb. 34**). Mit der in Abschnitt 3.5 beschriebenen Untersuchung von Zellpellet-Extrakten und Medienüberständen konnte diesem Cluster noch kein Metabolit zugewiesen werden. Zur Identifizierung des Metaboliten wurde deshalb folgende

Strategie gewählt: In den codierenden Abschnitt der NRPS-Komponente NpF3173 wurde mit Hilfe eines „Triparental Mating“ eine Kanamycin-Resistenzkassette eingeführt, so dass die Mutante *NpF3173::Kan^R* erhalten wurde. Zellpellet-Extrakt und Medienüberstand dieser Mutante wurden dann wiederum mittels HPLC-Auftrennung untersucht. Diese Ergebnisse wurden mit den Chromatogrammen der Wildtyp-Proben von *N. punctiforme* verglichen, um den Metaboliten zu identifizieren.

3.6.3.1 Überprüfung der Mutante *NpF3173::Kan^R*

Um zu testen, ob Exkonjuganten-Kolonien der „Triparental Mating“-Prozedur das Genfragment durch ein doppeltes Rekombinationsereignis in das Genom integriert hatten, wurde zunächst ihr Wachstum auf Saccharose-haltigen und Erythromycin-haltigen Kulturplatten überprüft. Da der für den Transfer verwendete Vektor pRL271 das *sacB*-Gen und eine Erythromycin-Resistenzkassette enthält, kamen nur solche Exkonjuganten-Kolonien in Betracht, die auf den Saccharose-haltigen Platten wachsen konnten (positive Selektion), nicht jedoch auf den mit Erythromycin versetzten Platten (negative Selektion). Zur weiteren Sicherstellung wurde genomische DNA aus den potenziellen *NpF3173::Kan^R*-Mutanten isoliert und mit Hilfe einer PCR die Lage der Kanamycin-Resistenzkassette überprüft. Als Primer wurden NpF3173KOcheck1 und NpF3173KOcheck2 jeweils in Kombination mit den Primern KM1800-FWD und KM375-REV verwendet, mit denen aus der integrierten Kanamycin-Resistenzkassette herausgelesen wurde. Da die beiden erstgenannten Primer in Abschnitten der benachbarten Gene liegen, können in dieser Kontroll-PCR nur dann Fragmente von ca. 2000 bp Länge amplifiziert werden, wenn der Vektor nicht mit in das Genom integriert wurde.

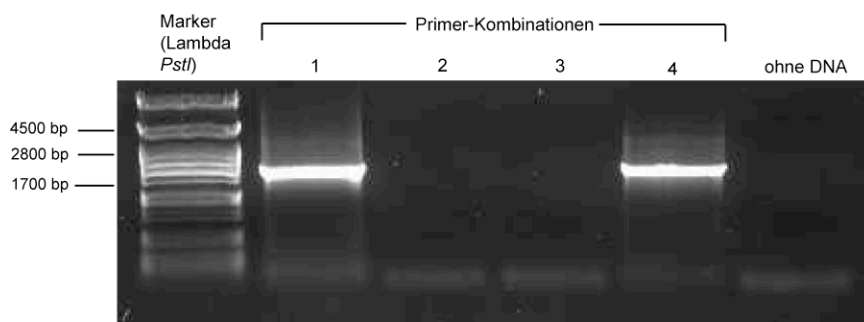


Abb. 48: Überprüfung der Mutante *NpF3173::Kan^R* mittels PCR.

Primer-Kombinationen: 1: NpF3173KOcheck1 + KM1800-FWD; 2: NpF3173KOcheck1 + KM375-REV; 3: NpF3173KOcheck2 + KM1800-FWD; 4: NpF3173KOcheck2 + KM375-REV

In der PCR konnten mit den passenden Primer-Kombinationen diejenigen Fragmentlängen amplifiziert werden, die bei Integration des DNA-Stücks ohne den Vektor zu erwarten sind. Somit sprachen alle Testverfahren dafür, dass es sich bei der ausgewählten Mutantenkolonie tatsächlich um eine *NpF3173::Kan^R*-Mutante handelte.

3.6.3.2 Auswirkungen der Knockout-Mutation

Aus Zellmaterial, das auf Kulturplatten angezogen worden war, wurde ein Proteinextrakt hergestellt und mit Extrakten aus zwei Wildtyp-Proben von *N. punctiforme* sowie einer Wildtyp-Probe von *Synechocystis* PCC 6803 verglichen. Dazu wurden jeweils 10 mg Gesamtprotein der vier Proben elektrophoretisch mittels SDS-PAGE aufgetrennt und mit Hilfe eines *Western Blotting* auf eine Transfermembran übertragen. Auf der Membran konnte dann die Existenz von A-Domänen mit Spezifität für die Aminosäure Cystein immunchemisch nachgewiesen werden. Als Primärantikörper wurde ein Antikörper aus Kaninchen eingesetzt, der eine hohe Spezifität für solche A-Domänen aufwies (zur Verfügung gestellt von Dr. Torsten Stachelhaus).

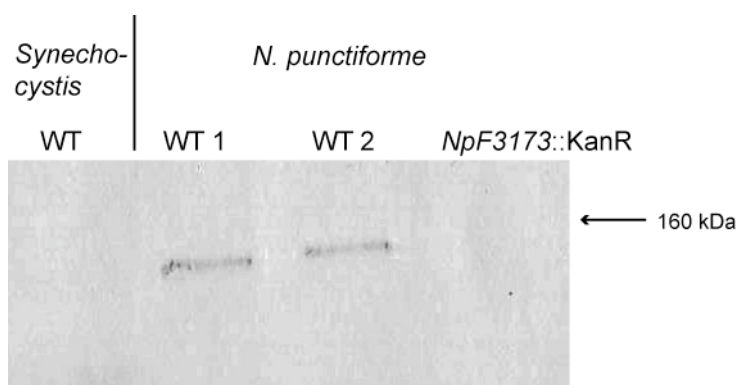


Abb. 49: Immunchemischer Nachweis von Cystein-aktivierenden A-Domänen in Zellextrakten

Jeweils 20 µg Gesamtproteinextrakte von Wildtyp-Proben von *N. punctiforme* und *Synechocystis* PCC 6803 sowie der *N. punctiforme*-Mutante *NpF3173::Kan^R* wurden mittels SDS-PAGE aufgetrennt und mit *Western Blotting* auf eine Membran transferiert. Die Inkubation erfolgte mit einem primären Antikörper gegen Cystein-aktivierende A-Domänen und einen sekundären Antikörper mit gekoppelter Alkalischer Phosphatase. Der Nachweis erfolgte colorimetrisch.

Das Ergebnis dieses vergleichenden Nachweises ist in **Abb. 49** dargestellt. Die Probe aus *Synechocystis* wurde als Negativprobe mitgeführt, da aus der Genomsequenz dieses Cyanobakteriums bekannt ist, dass dort keine NRPS-Proteine codiert sind. Erwartungsgemäß waren in dieser Spur keine Signale zu beobachten. In den Wildtyp-Extrakten von *N. punctiforme* konnte dagegen im richtigen Größenbereich jeweils eine gut definierte Bande angefärbt werden. Diese Bande fehlte in der Mutante *NpF3173::Kan^R*. Die NRPS-Komponente *NpF3173* wurde also in der Mutante nicht mehr zur Expression gebracht.

Der HPLC-Vergleich von Zellextrakten des Wildtyps und der Mutante, die wiederum mittels Festphasenextraktion aufkonzentriert waren, blieb unauffällig. Alle im Wildtyp sicher detektierbaren Peaks waren in der Mutante *NpF3173::Kan^R* ebenfalls zu finden (Daten nicht gezeigt). Ein wesentlicher Unterschied war hingegen beim Vergleich der Kulturüberstände sichtbar.

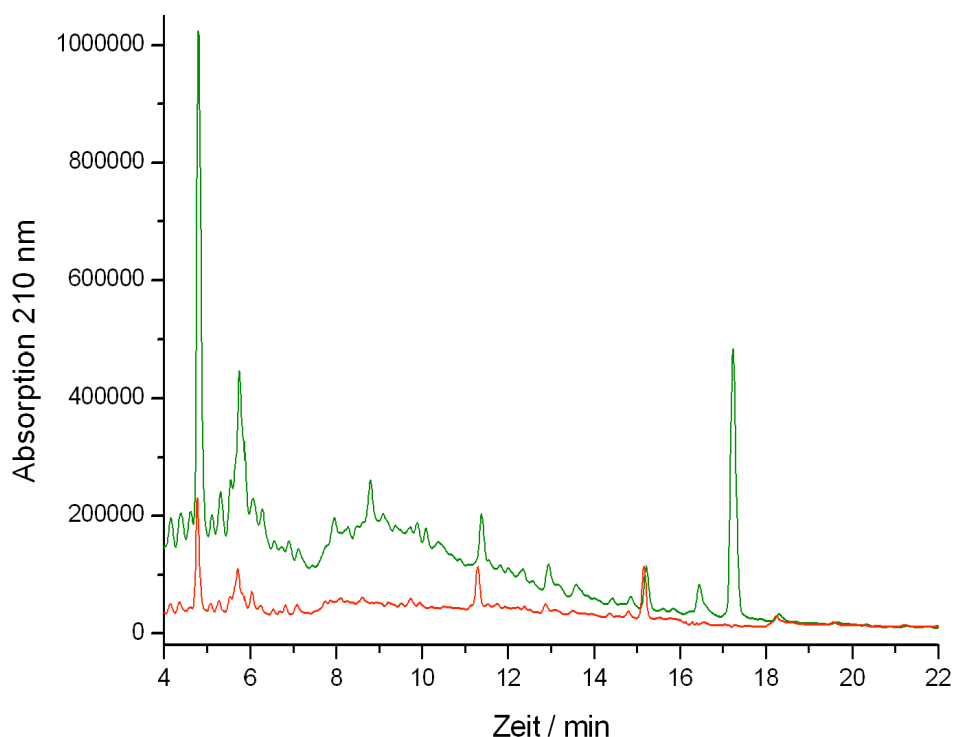


Abb. 50: Vergleich der HPLC-Läufe von Kulturüberständen des Wildtyps und der Mutante *NpF3173::Kan^R*.

Die aufkonzentrierten wässrigen Kulturüberstände wurden mit einem linearen Wasser/Acetonitril-Gradienten aufgetrennt. In der Abbildung sind die beiden HPLC-Läufe zur besseren Vergleichbarkeit übereinander gelagert. Das Chromatogramm der Wildtyp-Probe ist in grün dargestellt, das der Mutante in rot.

Wie aus **Abb. 50** zu erkennen ist, fehlt ein sehr prominenter Peak bei etwa 17 min ebenso wie ein kleinerer, etwas früher eluierender Begleit-Peak im Mediumüberstand der Mutante vollständig. Dieses Ergebnis konnte in mehreren unabhängigen Messungen bestätigt werden. Da zu diesem Peak im allgemeinen Screening (siehe Abschnitt 3.5) noch keine Daten hinsichtlich der Masse erhalten werden konnten, wurden größere Probenmengen, ausgehend von einem Liter Kultur, mittels HPLC-Läufen aufgetrennt und dabei der entsprechende Bereich fraktioniert. Die gesammelten Fraktionen wurden in einem ESI-Massenspektrometer analysiert (in Kooperation mit Katrin Schubert, AG Prof. R. Süßmuth, TU Berlin). Im fraglichen Bereich konnte eine Substanz mit $m/z=252$ detektiert werden. Andere Massen wurden nicht gefunden, so dass davon ausgegangen werden kann, dass der Peak auf eine einzelne Verbindung zurück zu führen ist.

Auf diesen Ergebnissen aufbauend, wurde die in Frage stehende Fraktion mittels einer Fouriertransformation-Ionencyclotronresonanz-Massenspektrometrie (FT-ICR-MS) untersucht (in Kooperation mit dem Institut für Organische Chemie der Eberhard-Karls-Universität Tübingen). Das

Messverfahren der FT-ICR-MS erlaubt sehr exakte Bestimmungen der Molekülmasse, so dass unter Berücksichtigung der Isotopenverteilungen der Elemente alle praktisch relevanten Summenformeln der Verbindung berechnet werden können. Die Messung ergab, dass die einzige chemisch sinnvolle Summenformel für die unbekannte Verbindung C₁₅H₂₄O₃ lautete. Insbesondere konnte anhand der FT-ICR-MS-Messung mit Sicherheit ausgeschlossen werden, dass die Verbindung ein Schwefel-Atom enthält, was jedoch nach Einbau eines Cystein-Restes, der in der Verbindung z. B. als Thiazolidin-Ring vorliegt, zu erwarten gewesen wäre.

Zusammengefasst konnte also die *N. punctiforme*-Mutante *NpF3173::Kan^R* erhalten werden, welche die NRPS-Komponente des Genclusters VII nicht mehr zu exprimieren vermag. Durch das Fehlen eines Signals im HPLC-Lauf des Kulturüberstands dieser Mutante konnte eine Verbindung ermittelt werden, die möglicherweise diesem Gencluster VII zugeordnet werden kann.

4 Diskussion

4.1 Evolutionäre Einordnung der Polyketidsynthasen

Fragt man nach den evolutionären Entwicklungslinien von Enzymen des Sekundärmetabolismus, so können in Abhängigkeit des Untersuchungszusammenhangs mehrere Stufen unterschieden werden. Dies ist in **Abb. 51** für die bakteriellen Polyketidsynthasen (PKS) graphisch veranschaulicht. Zunächst bietet sich auf der obersten Stufe ein Vergleich mit verwandten Enzymen des Primärstoffwechsels an. Das Verwandtschaftsverhältnis kann sich in diesem Zusammenhang auf zwei unterschiedliche Aspekte beziehen: zum einen Ähnlichkeiten in den Sequenzen, die auf einen gemeinsamen Vorläufer und somit auf das Vorliegen von Homologie schließen lassen, zum anderen Übereinstimmungen im enzymatischen Mechanismus bezüglich der Substrate und ihrer Umsetzung. Zu beachten ist dabei im Allgemeinen, dass der letztgenannte funktionelle Aspekt nicht zwangsläufig auf eine evolutionär gemeinsame Herkunft schließen lässt, während umgekehrt homologe Proteine recht unterschiedliche Funktionen besitzen können (Theißen 2002).

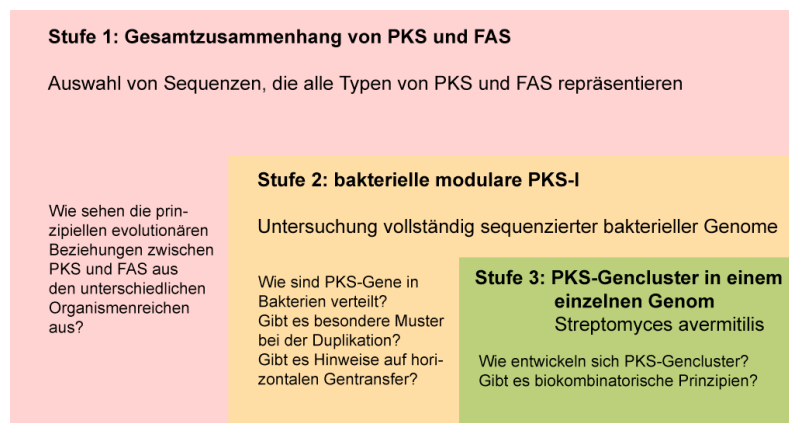


Abb. 51: Die verschiedenen Stufen bei der Untersuchung evolutionärer Entwicklungslinien

Im Fall der PKS-I ist dies unproblematisch, da ihre Verwandtschaft zu den FAS beide genannten Aspekte umfasst. Sehr frühzeitig wurde erkannt, dass die Biosynthese von Polyketid-Verbindungen enzymatische Aktivitäten erfordert, die im Anabolismus von Fettsäuren vorkommen (zusammengefasst bei (Birch 1967)). Die Wurzeln dieser Hypothese reichen bis an den Beginn des 20. Jahrhunderts zurück (Collie 1907). Neben dieser eindeutigen funktionellen Verwandtschaft sind auch die Sequenzhomologien der verschiedenen Domärentypen zweifelsfrei belegt, seitdem die ersten Teilsequenzen von PKS-Genen aufgeklärt werden konnten (Hopwood und Sherman 1990; Donadio *et al.* 1991; Bevit *et al.* 1992; Donadio und Katz 1992). Durch eine phylogenetische Untersuchung repräsentativer Sequenzen von FAS und PKS aus verschiedenen Organismenreichen sind Rückschlüsse auf den evolutionären Gesamtzusammenhang

möglich. Dieser wird in Abschnitt 4.1.1 diskutiert. Den erstaunlich engen evolutionären Beziehungen zwischen eukaryotischen FAS, insbesondere denjenigen aus Tieren, und modularen bakteriellen PKS des Typs I ist der Abschnitt 4.1.2 gewidmet. PKS des Typs III werden aufgrund ihrer eigenständigen Position getrennt von den anderen Typen in Abschnitt 4.1.3 behandelt.

Engt man den Untersuchungsgegenstand ein, gelangt man zur mittleren Stufe 2, die sich mit den modular aufgebauten PKS-I aus Bakterien beschäftigt. Bei der Analyse eines Datensatzes, der die Auswertung einer großen Zahl von vollständig sequenzierten bakteriellen Genomen in sich vereinigt, stehen im Vordergrund Fragen nach der Verteilung von PKS-I-Genen in den diversen Bakteriengruppen und nach Duplikationsmustern und eventuellen Hinweisen auf Prozesse horizontalen Gentransfers. Die Diskussion dieser Aspekte erfolgt in Abschnitt 4.2.

Eine weitergehende Fokussierung führt schließlich zur Stufe 3, auf der versucht wird, die evolutionäre Entwicklung von einigen wenigen PKS-Genclustern gewissermaßen zu sezieren, um sich ein Bild davon zu machen, welche besonderen Mechanismen hier zum Tragen kommen und ob es eine natürliche Variante der im Labor praktizierten biokombinatorischen Ansätze gibt. Die in Abschnitt 3.3 dargelegten Ergebnisse zur intragenomischen Evolution dieser Cluster am Beispiel des Modellorganismus *Streptomyces avermitilis* werden in Abschnitt 4.3 diskutiert.

4.1.1 Phylogenetische Zusammenhänge zwischen PKS-I, PKS-II und FAS

Der in **Abb. 8** im Ergebnisteil (Abschnitt 3.1.1) dargestellte phylogenetische Baum ist in **Abb. 52** nochmals in vereinfachter und übersichtlicherer Form skizziert. Aus der in der Einleitung (Abschnitt 1.2.1.2) in **Abb. 4** gegebenen vergleichenden Übersicht zu den verschiedenen Typen von FAS und PKS geht hervor, dass bei der Interpretation der evolutionären Verhältnisse zwei Gesichtspunkten besondere Beachtung geschenkt werden muss. Erstens ist dies der Unterschied zwischen distinkten Systemen einerseits, deren enzymatische Aktivitäten auf verschiedenen Proteinen lokalisiert sind, und fusionierten Systemen andererseits, in denen diese Aktivitäten durch hintereinander liegende Domänen realisiert werden. Zweitens ist im Hinblick auf die „Arbeitsweise“ zwischen iterativen und nicht-iterativen Biosynthesemechanismen zu unterscheiden.

Die Hinzunahme von FabH und FabF aus Archaea und Eubakterien diene dazu, mit Hilfe von „Außengruppen“ (englisch: *outgroups*) eine Verwurzelung des Baums zu ermöglichen und damit die Entwicklungsrichtung innerhalb der Phylogenie sichtbar zu machen. Geeignete *outgroups* sind dadurch gekennzeichnet, dass sie die relativ zu den anderen untersuchten operationalen taxonomischen Einheiten (OTU; von englisch: *operational taxonomic units*) am weitesten entfernten Homolge darstellen (Vandamme 2003). Üblicherweise sollten Indizien vorliegen, die eine Aufspaltung dieser *outgroup*-Sequenzen vor den Zeitpunkten der Aufspaltungen der anderen OTU nahe legen (Graur und Li 1999). Die Wahl von Sequenzen der Proteine FabH und FabF geschah also unter der Annahme, dass die Verzweigungspunkte 1 und 2 in **Abb. 52** diesen Anforderungen genügen.

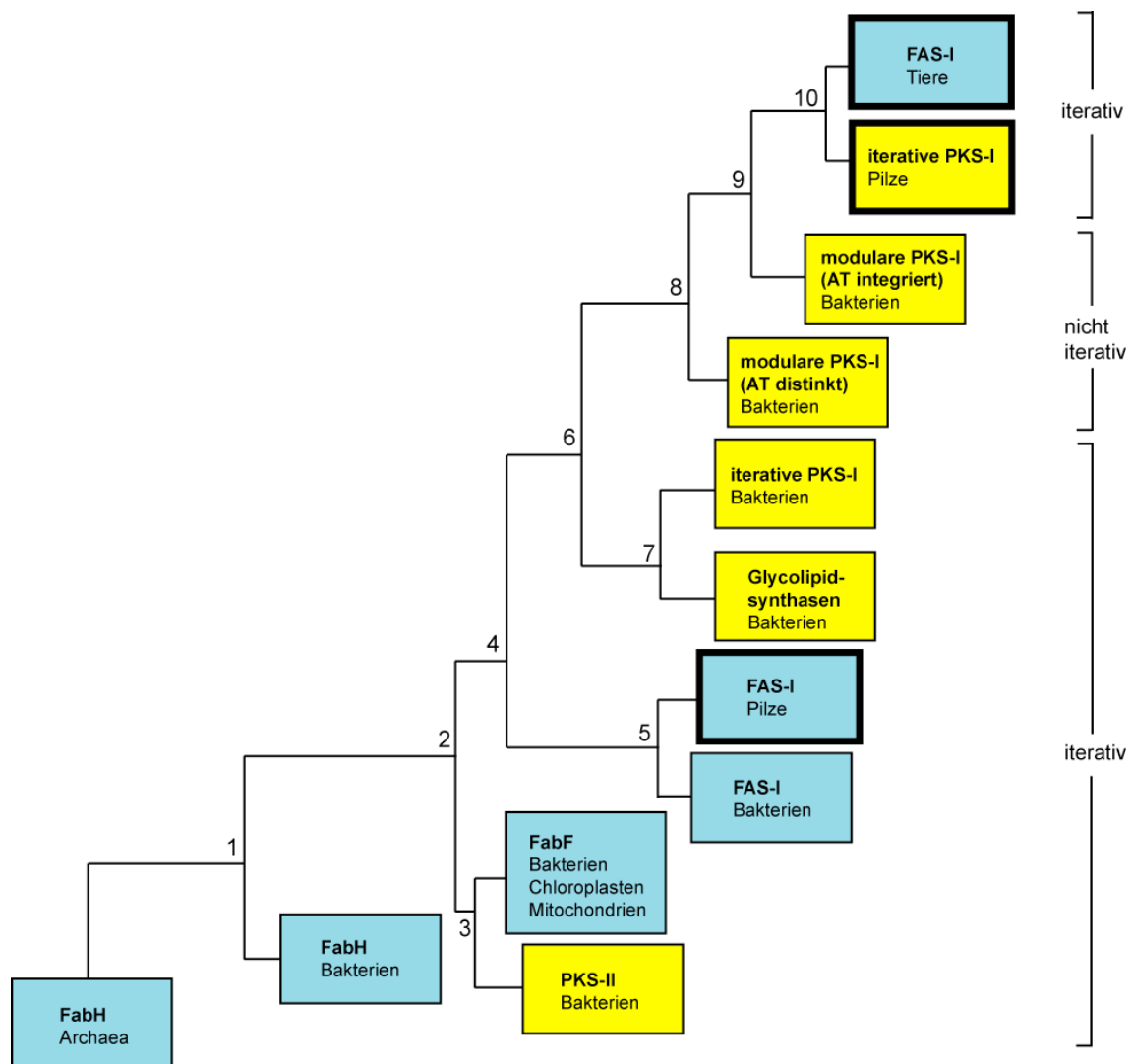


Abb. 52: Phylogenetische Zusammenhänge zwischen PKS und FAS.

Die Skizze basiert auf dem detaillierten phylogenetischen Baum aus **Abb. 8** in Abschnitt 3.1.1 des Ergebnisteils. Alle Formen von PKS sind gelb unterlegt, FAS-Typen blau. Proteine aus Eukaryoten sind mit einem dicken Rahmen gekennzeichnet. Die Zahlen dienen zur Nummerierung der Verzweigungspunkte, auf die im Text Bezug genommen wird.

Diese Hypothese wird durch folgende Indizien gestützt: (1) Prokaryoten existierten lange Zeit vor dem Auftreten eukaryotischer Vorläufer. Diese Aussage fußt auf einer Vielzahl von phylogenetischen Untersuchungen, die sowohl auf der Basis von ribosomalen RNA-Sequenzen als auch Sequenzen hoch konservierter Proteine erstellt wurden (Fox *et al.* 1980; Iwabe *et al.* 1989; Woese, Kandler und Wheelis 1990; Doolittle 1999; Battistuzzi, Feijao und Hedges 2004). (2) Die Isoenzym-Formen FabH und FabF finden sich allgemein bei allen Bakterien (Marrakchi, Zhang und Rock 2002), weshalb von einer frühen Duplikation und Spezialisierung ausgegangen werden kann. Somit kann von der Zulässigkeit dieser *outgroups* ausgegangen werden.

Besondere Erwähnung verdient der Umstand, dass sich in den ausgewählten eukaryotischen Genomen grundsätzlich Komponenten des Bakterien-typischen FAS-II finden, so etwa die FabF-Proteine. Der Ursprung dieser in den Mitochondrien lokalisierten Proteine ist auf die endosymbiontische Aquisition eines Vorläufers der rezenten α -Proteobakterien zurückzuführen (Yang *et al.* 1985; Cedergren *et al.* 1988) und liegt daher in direkter Nähe zur FabF-Sequenz aus *Agrobacterium tumefaciens* (siehe **Abb. 8**). Es gibt in Reihe von Hinweisen, dass dieses mitochondriale FAS-System wichtig für die Reparatur von Phospholipiden der Mitochondrien-Membrane (Schneider *et al.* 1997) und für die Bereitstellung von Octanoyl-ACP als Vorläufer in der Biosynthese von Liponsäure und anderen langkettigen Fettsäuren ist (Yasuno, von Wettstein-Knowles und Wada 2004; Zhang *et al.* 2005).

4.1.1.1 Phylogenetische Position der PKS-II

Bakterielle PKS-II, die aus distinkten Proteinen aufgebaut sind, stellen sich als Schwestergruppe von FabF dar (Verzweigungspunkt 3 in **Abb. 52**) und liegen weit entfernt von allen PKS des Typs I mit ihren fusionierten Domänen. Sie stellen also eine Eigenentwicklung dar, die keine engeren Bezugspunkte zu den anderen PKS aufweist. Dies scheint sich auch in ihrer taxonomisch sehr engen Verbreitung zu spiegeln, denn lange Zeit ging man davon aus, dass PKS-II ausschließlich in Actinomyceten vorkommen (Rawlings 1999; Staunton und Weissman 2001). Allerdings wurde kürzlich ein Enzym aus dem Myxobakterium *Stigmatella aurantiaca* beschrieben, das sehr große Ähnlichkeiten mit PKS-II aufweist (Sandmann *et al.* 2007). Es muss abgewartet werden, ob mit der Verfügbarkeit weiterer vollständiger Genomsequenzen die Zahl an PKS-II-Genen in anderen Bakteriengruppen als den Actinomyceten weiter ansteigt und sich dadurch Änderungen in der Auffassung ergeben, es handele sich um eine monophyletische und recht klar abgegrenzte Sonderentwicklung.

4.1.1.2 Verwandtschaft zwischen bakteriellen FAS-I und FAS aus Pilzen

Im nächsten Cladus, der sich aus dem Verzweigungspunkt 4 (siehe **Abb. 52**) ergibt, sind zwei über den Punkt 5 verknüpfte Subcladi vereint: die FAS aus Pilzen und diejenigen FAS-I, die aus der so genannten CMN-Gruppe (CMN: Corynebakterien, Mycobakterien und Nocardien) der Bakterien stammen. Es handelt sich um die beiden ersten Gruppen im Baum, die fusionierte, auf einer Polypeptid-Kette lokalisierte Aktivitäten aufweisen. Wenn man von „fusionierten“ Aktivitäten spricht, impliziert dies bereits das Bild einer Proteinevolution durch Genfusionen. Tatsächlich liegt es aufgrund der oben genannten Argumente nahe, davon auszugehen, dass die FAS-II mit ihren Einzelproteinen die ursprüngliche Form der Fettsäuresynthese darstellt. Im Fall von schrittweisen Genfusionen wären strukturell und funktionell relativ unabhängige Domänen auf der multifunktionellen Polypeptid-Kette zu erwarten. Hinweise dafür lagen bereits frühzeitig durch genetische Experimente an der FAS aus Hefe vor (Schweizer, Werkmeister und Jain 1978) und wurden kürzlich durch die Aufklärung der Kristallstruktur der FAS aus dem Schlauchpilz *Thermomyces lanuginosus* bestätigt (Jenni *et al.* 2006). Auf den Aspekt der Genfusion und seine evolutionäre Bedeutung wird weiter unten im Rahmen der speziellen Diskussion der FAS-Evolution genauer eingegangen (siehe Abschnitt 4.1.2).

Die erstaunliche evolutionäre Nähe der FAS aus Pilzen und Bakterien der CMN-Gruppe ergab sich als interessanter Nebenaspekt der intendierten Untersuchung der phylogenetischen Verhältnisse zwischen PKS und FAS. Die Positionierung im Baum als Schwestergruppen, die von einem direkten gemeinsamen Vorläufer abstammen, zeigt deutlich die enge Verwandtschaft zwischen beiden Synthese-Systemen bezüglich ihrer KS-Domänen. Exakt die gleichen Konstellationen wurden für die AT- und KR-Domänen in entsprechenden Datensätzen gefunden, die alle Organismenreiche abdecken (siehe **Abb. 9**). Für die DH- und ER-Domänen gilt, dass sie jeweils große Sequenzübereinstimmungen zwischen Pilzen und Vertretern der CMN-Gruppe zeigen, jedoch völlig anderen Proteinfamilien angehören als die entsprechenden Proteine und Domänen der anderen Prokaryoten bzw. Tieren (Fox und Lynen 1980; Jenni *et al.* 2006). So verwenden beide eine ER-Domäne, die FMN als Cofaktor enthält, während die ER-Aktivitäten der Bakterien und Tiere keinen Cofaktor benötigen, sondern NADPH direkt als Reduktionsmittel benutzen. Schließlich fällt auch die völlige Übereinstimmung in der Domänenanordnung (Meurer *et al.* 1992; Fernandes und Kolattukudy 1996) ins Auge, wenn man die beiden Proteine aus Pilzen in passender Reihenfolge hintereinander setzt, wie in der Einleitung gezeigt (siehe **Abb. 4** in Abschnitt 1.2.1.2). Nimmt man alle genannten Argumente zusammen, so bleibt kein vernünftiger Zweifel darüber, dass es sich bei den eukaryotischen Pilz-FAS und den prokaryotischen FAS-I der CMN-Gruppe um evolutionär sehr nahe verwandte Proteine handelt, die von einer direkten gemeinsamen Entwicklungslinie unmittelbar ableitbar sind. Da Pilze und Actinobakterien, zu denen die Vertreter der CMN-Gruppe gehören, offenkundig evolutionär weit entfernte Taxa darstellen, muss die Möglichkeit eines horizontalen Gentransfers (HGT) erwogen werden. Dieser Vorschlag gewinnt an weiterem Gewicht, wenn man bedenkt, wie unwahrscheinlich es ist, dass sich die beiden FAS-Systeme mit all den aufgezählten Übereinstimmungen völlig unabhängig voneinander entwickelt haben könnten. Eine Erklärung der Verhältnisse als Konvergenz-Phänomen scheidet gewiss aus.

Diesem Umstand wird in der Literatur erstaunlich wenig Aufmerksamkeit geschenkt, obwohl speziell der Fettsäure-Metabolismus in den diversen Mycobakterien-Stämmen seit langem intensiv beforscht wird, da eine Gruppe von langkettigen Fettsäuren, die so genannten Mycolsäuren, bei der Infektion beispielsweise durch die humanpathogenen Species *Mycobacterium tuberculosis* und *Mycobacterium leprae* eine wichtige Schutzfunktion für den Krankheitserreger haben. Der Fettsäure-Stoffwechsel von Mycobakterien ist daher grundsätzlich bedeutsam als möglicher Angriffspunkt für therapeutische Wirkstoffe (Barry *et al.* 1998). Seit der vollständigen Sequenzierung der Genome zweier Stämme von *M. tuberculosis* (Cole *et al.* 1998; Valway *et al.* 1998) ist bekannt, dass Mycobakterien eine im Vergleich zu anderen Bakteriengruppen sehr große Zahl an Genen besitzen, die in den Stoffwechsel von Fettsäuren involviert sind. So wird die Zahl bei *M. tuberculosis* auf etwa 250 geschätzt, während *Escherichia coli* mit seinem ähnlich großen Genom nur ungefähr 50 enthält (Cole *et al.* 1998). Eine der Besonderheiten des Fettsäure-Stoffwechsels innerhalb der CMN-Gruppe liegt darin, dass die für Bakterien typischen FAS-II-Enzyme nicht zur *de novo*-Synthese von Fettsäuren befähigt sind. Diese erfolgt ausschließlich am FAS vom Typ I bis zu einer Länge von 14 oder 16 C-Atomen und wird dann gegebenenfalls zur weiteren Verlängerung an die FAS-II transferiert (Schweizer und Hofmann 2004).

Eine Studie, deren Ziel explizit im Auffinden mycobakterieller Gene eukaryotischen Ursprungs bestand, kam zu dem Schluss, dass es im Genom von *M. tuberculosis* insgesamt 19 Verdachtskandidaten gebe (Gamielidien, Ptitsyn und Hide 2002). Die Gene der FAS-I wurden jedoch nicht als Kandidaten eines

HGT-Ereignisses identifiziert. Dies erklärt sich vermutlich aus einem zu strikten Kriterium bei der Auswahl möglicher eukaryotischer Vorläufer-Proteine. Um in den Kandidatenkreis aufgenommen zu werden, musste das eukaryotische Protein bei der BLAST-Analyse einen E-Wert aufweisen, der um mindestens zehn Größenordnungen höher war als der Wert für das nächstverwandte prokaryotische Protein. Aus demselben Grund blieben die FAS-I-Gene auch bei einer früheren Arbeit unerkannt, die so genannte „Interkingdom“-Genfusionen aufzudecken versuchte und mit einer ähnlichen Methodik der Auswahl nach bestimmten BLAST-Kriterien vorging (Wolf, Kondrashov und Koonin 2000). Unter „Interkingdom“-Genfusionen werden dabei solche Proteine verstanden, die aus Domänen bestehen, die aus unterschiedlichen Organismenreichen stammen. Interessanterweise konstatieren die Autoren dieser Publikation, dass die mit Abstand höchste Zahl an derartigen Genfusionen im Genom von *M. tuberculosis* gefunden werden konnte, wobei diese am leichtesten durch HGT der entsprechenden Domänen aus Eukaryoten und nachfolgende Fusion mit Genen des ursprünglichen mycobakteriellen Bestands zu erklären sind (Wolf, Kondrashov und Koonin 2000). Schließlich wurde in einer weiteren Studie der Frage nachgegangen, inwieweit HGT, Genduplikationen und Adaptationen zur Evolution der Enzyme des Fettsäure-Metabolismus in *M. tuberculosis* beigetragen haben (Kinsella *et al.* 2003). Es stellte sich heraus, dass in fünf der insgesamt acht untersuchten Genfamilien deutliche Anzeichen für HGT-Ereignisse vorliegen, insbesondere auffällig stark ausgeprägte phylogenetische Verwandtschaftsverhältnisse mit α -Proteobakterien. Allerdings war diese Studie auf einen Vergleich mit ausschließlich bakteriellen Genomen beschränkt, weshalb keine Ergebnisse bezüglich HGT von Eukaryoten zu erwarten waren. Dennoch lässt sich aus den drei genannten Publikationen ein aufschlussreicher Gedanke entnehmen: Die Mycobakterien gelten gemeinhin als „hochverdächtig“, wenn es um die Frage möglicher HGT-Ereignisse geht, und in ihren Genomen fanden sich bislang zahlreiche Hinweise auf solche Prozesse. Eine spätere Arbeit (Kinsella und McInerney 2003) relativierte zwar die von Gamielien *et al.* gefundene Zahl von 19 Genen eukaryotischen Ursprungs durch alternative Erklärungen, jedoch bleibt im Gesamtbild der vorgelegten Forschungsarbeiten die Aussage weiterhin gültig.

Zusammengenommen lässt sich also folgender evolutionärer Ablauf konstruieren: Die heute in Pilzen zu findende FAS mit ihrer charakteristischen Domänenanordnung ist ein Fusionsprodukt ursprünglich eigenständiger Proteine, wie man sie in rezenten Bakterien findet. Dieses fusionierte Biosynthese-System wurde durch einen HGT auf einen Vorläufer der heutigen prokaryotischen CMN-Gruppe übertragen, wobei es in einzelnen Gruppen wie den Mycobakterien zu einem weiteren Fusionsereignis kam, wodurch aus den zwei Einheiten der Pilze (α -Untereinheit und β -Untereinheit, häufig auch als FAS1 und FAS2 bezeichnet) ein noch größeres Protein entstand. Dieses Szenario ist notgedrungen zu einem gewissen Grad spekulativ, erklärt jedoch die vorliegenden Umstände in ihrer Gesamtheit am besten. Auf den wichtigen Aspekt, dass die Genfusionsereignisse, die zur Ausbildung der heutigen Pilz-FAS führten, mit hoher Wahrscheinlichkeit unabhängig von den zur tierischen FAS führenden Genfusionen waren, wird weiter unten in Abschnitt 4.1.2 eingegangen. Dort wird die Evolution der FAS nochmals einer Gesamtschau unterzogen.

Grundsätzlich stellt sich die Frage, warum im Bereich der Prokaryoten praktisch ausschließlich FAS vom Typ II zu finden sind, während die fusionierten Varianten des Typs I mit Ausnahme der CMN-Gruppe eukaryotischen Organismen vorbehalten zu sein scheinen. Ein Multienzymkomplex weist offensichtliche kinetische Vorteile gegenüber einem „System“ aus nicht assoziierten Enzymen auf, wobei allerdings davon

auszugehen ist, dass die distinkten Proteine des bakteriellen FAS-II in dieser Form organisiert sind und somit effizient kooperieren (Rock und Jackowski 2002). Der besondere Vorteil einer Konfiguration, in der die einzelnen Komponenten auf einer gemeinsamen Polypeptid-Kette lokalisiert sind, besteht nach herrschender Meinung darin, dass die benötigten Aktivitäten automatisch in der richtigen Stöchiometrie produziert werden und dass die Assemblierung des Gesamtkomplexes erleichtert wird. Dies wird jedoch mit einem Nachteil erkauft, der für den Organismus äußerst schädlich sein kann. Die Wahrscheinlichkeit, dass während der ribosomalen Synthese von Proteinen ein die Funktion beeinträchtigender Fehler unterläuft, nimmt selbstverständlich mit der Länge des Proteins zu. Da die Genauigkeit der prokaryotischen Translation geringer ist als die in Eukaryoten, mag es für Bakterien vorteilhaft sein, für eine so essenzielle Aufgabe wie die Fettsäuresynthese den ursprünglichen Modus beizubehalten. Dennoch ist anzunehmen, dass es in Prokaryoten nach Duplikationen zu diversen Genfusionen kam, die zu fusionierten Proteinen des Typs I führten. Dies folgt allein schon aus der Existenz einer Reihe von spezifisch bakteriellen fusionierten Systemen, die in Abb. 52 oberhalb des Verzweigungspunktes 4 zu finden sind.

4.1.1.3 Bakterielle iterative PKS-I und Glycolipid-Synthasen

In der Phylogenie folgt ein großes Cluster ab dem Verzweigungspunkt 6 (**Abb. 52**), das sich aus bakteriellen iterativen PKS-I und Glycolipidsynthasen zusammensetzt, die wiederum einen nahen gemeinsamen Vorläufer (Verzweigungspunkt 7) haben. Diese beiden Biosynthese-Systeme können als eigener Entwicklungszweig interpretiert werden, der sich durch folgende Eigenschaften charakterisieren lässt: (1) Die Enzymsysteme bestehen aus fusionierten Einheiten, die als gut abgrenzte Domänen in Erscheinung treten, (2) sie arbeiten iterativ, und (3) sie bedienen „Sondersynthesewege“, wodurch sich ihre für Enzyme des Sekundärmetabolismus typische spezifische Verbreitung erklärt. Daraus mögen sich auch ihre im Vergleich mit anderen PKS-I erkennbaren strukturellen und funktionellen Besonderheiten erklären, z. B. das Vorliegen multipler ACP-Domänen (siehe **Abb. 4** in Abschnitt 1.2.1.2). In diesem Bereich finden sich die iterativen PKS-I aus *Photobacterium profundum*, einem zur Familie *Vibrionaceae* gehörenden γ -Proteobakterium, und aus *Shewanella oneidensis*, das ebenfalls in die γ -Proteobakterien und dort in die Familie *Shewanellaceae* eingeordnet wird. Diese iterativen PKS-I sind für die Biosynthese von mehrfach ungesättigten Fettsäuren (PUFA, *polyunsaturated fatty acid*) zuständig. PUFA sind seit langer Zeit bekannt als wichtige Membranbestandteile von Zellen, die in bestimmten Geweben in erheblicher Menge vorkommen. So sind sie etwa in hohem Anteil in der Retina und im Gehirn zu finden (Lauritzen *et al.* 2001). Darüber hinaus sind sie die Vorläufer wichtiger Gewebshormone wie Prostaglandine, Leukotriene und Thromboxane (Heller und Koch 1998). Bis vor etwa zehn Jahren ging man davon aus, dass PUFA ausschließlich in eukaryotischen Organismen vorkommen. Ab Mitte der 90er Jahre des 20. Jahrhunderts wurden jedoch mehrere Arbeiten publiziert, die das Vorkommen dieser Verbindungsklasse in einer Reihe von bakteriellen Species beschrieben (Yazawa 1996; Watanabe *et al.* 1997; Nichols *et al.* 1999; Russell und Nichols 1999), wobei das gemeinsame Merkmal dieser Species ihre Psychrophilie ist. Vermutlich besteht die besondere physiologische Rolle der PUFA in dieser Bakteriengruppe darin, unter niedrigen Temperaturen die notwendigen Fließeigenschaften der Zellmembran zu gewährleisten. Der genaue Mechanismus der PUFA-Biosynthese ist zwar noch nicht aufgeklärt, allerdings sind in der Literatur

Modelle vorgeschlagen worden (Kaulmann und Hertweck 2002; Wallis, Watts und Browse 2002), die insbesondere davon ausgehen, dass im Gegensatz zur eukaryotischen PUFA-Biosynthese keine spezifischen Desaturasen notwendig sind, sondern statt dessen eine bifunktionelle Dehydratase/2-*trans*,3-*cis*-Isomerase zum Einsatz kommt. Ungeklärt bleibt bislang auch der Grund für das Vorliegen multipler ACP-Domänen als durchgehend erkennbares Merkmal dieser Enzymgruppe.

Entsprechendes gilt für die ebenfalls in diesem phylogenetischen Cluster lokalisierten bakteriellen Glycolipid-Synthasen, die im verwendeten Datensatz vornehmlich aus Cyanobakterien stammen. Es existieren zwar einige Veröffentlichungen, in denen auf Basis genetischer Experimente die Beteiligung dieser PKS an der Biosynthese von Glycolipiden im Zusammenhang mit der Differenzierung von Heterocysten demonstriert werden konnte, jedoch liegen auch für diese Enzyme keine näheren biochemischen Daten vor, aus denen sich Schlussfolgerungen über den Ablauf der Biosynthese ziehen ließen. Die Struktur der Produkte legt jedenfalls einen iterativen Mechanismus nahe. Wie die Substratspezifitäten und Zykluslängen während der Synthese gesteuert werden, ist hingegen unbekannt.

Zusammenfassend kann festgestellt werden, dass dieses Cluster eine Reihe von speziellen Enzymen enthält, deren Biochemie noch weitgehend unerforscht ist, so dass nur wenig über ihre Eigenschaften ausgesagt werden kann, abgesehen davon, dass in allen Fällen von Iterativität ausgegangen werden muss.

4.1.1.4 Phylogenetische Position der Varianten modular aufgebauter PKS-I

Ab dem Verzweigungspunkt 6 treten im phylogenetischen Baum (**Abb. 52**) modulare PKS-I aus Bakterien auf. Dieser Zweig markiert nicht nur den Übergang zu einer modularen Bauweise, sondern auch zu einem nicht mehr iterativen Syntheseprinzip, denn jedes Modul baut nur einen weiteren Baustein in die wachsende Polyketid-Kette ein, wird also nicht mehrfach benutzt, wie bei den ursprünglichen iterativen Systemen üblich. Diese Nicht-Iterativität ist ein entscheidendes Kennzeichen der modularen PKS-I. Sie scheint dann im weiteren Verlauf der Evolution mit Erreichen der FAS aus Tieren und den PKS aus Pilzen wieder aufgehoben zu werden (siehe hierzu die ausführliche Diskussion in Abschnitt 4.1.2). Die einzige Ausnahme von der Nicht-Iterativität im verwendeten Datensatz bildet das Modul AurA aus dem Actinobakterium *Streptomyces thioluteus*, das im Verlauf der Aureothin-Biosynthese zweimal hintereinander verwendet wird (He und Hertweck 2003). Der dahinter stehende Mechanismus ist allerdings noch nicht geklärt. Solange nicht mehr Enzyme mit dieser besonderen Eigenschaft gefunden worden und entsprechende Sequenzen veröffentlicht sind, lässt sich keine Aussage darüber treffen, ob es sich eventuell um ein eigenes Subcluster handeln könnte.

Ein weiterer mit der Modularisierung einher gehender Trend ist die in manchen Bakteriengruppen festzustellende enorme Zuspitzung im Hinblick auf Genfusionen und die daraus resultierenden riesigen Proteine, die in der Literatur zur Etablierung des Begriffs „Megasyntasen“ geführt haben (Weissman 2004). So befinden sich beispielsweise in Streptomyceten bis zu fünf komplette Module auf einer einzigen Polypeptid-Kette (vgl. **Tab. 5** in Abschnitt 3.2.1).

Über den Verzweigungspunkt 8 erscheinen zunächst die modularen PKS-I, die keine integrierten AT-Domänen enthalten, sondern durch so genannte *trans*-AT funktionell komplementiert werden (siehe

hierzu die Einleitung dieser Arbeit, Abschnitt 1.2.2.1.3). Hierbei scheint es sich um eine Sonderentwicklung zu handeln, wenn man bedenkt, dass alle fusionierten Systeme der vorherigen Zweige des phylogenetischen Baums AT-Domänen enthalten und dies auch durchweg für die oberhalb von Verzweigungspunkt 8 liegenden Proteine gilt. Im Vergleich zur Vielzahl an charakterisierten oder in Genomen identifizierten und bislang uncharakterisierten modularen PKS-I mit *cis*-ständiger AT-Domäne nimmt sich die Anzahl der bekannten „AT-losen“ modularen PKS-I recht bescheiden aus. Aus phylogenetischer Sicht auffällig ist diese Gruppe insofern, als innerhalb von ihr im Gegensatz zu den PKS-I mit integrierten AT-Domänen keinerlei Species-spezifischen Untergruppierungen auszumachen sind (Piel *et al.* 2004). Die PKS-I-Systeme mit *trans*-ständiger AT-Domäne sind in zweierlei Hinsicht für biokombinatorische Experimente von besonderem Interesse: Zum einen lässt sich, wie bereits gezeigt werden konnte, die Produktivität des Systems durch Überexpression der AT-Aktivität deutlich steigern (Cheng, Tang und Shen 2003), zum anderen scheint sich eine relativ einfache Möglichkeit zur Änderung der Spezifität für die neu einzubauenden Einheiten zu eröffnen. Abgesehen von diesen praktisch-experimentellen Aspekten kann darüber spekuliert werden, worin die biologischen Vorteile einer solchen Anordnung bestehen könnten. Augenfällig wäre ein gewissermaßen „ökonomischer“ Aspekt, denn wenn ein und dasselbe Enzym für die Beladung sorgt und in *trans* bereitgestellt wird, spart sich der Organismus offensichtlich einen bestimmten Codierungsaufwand in seinem Genom. Relevanter könnte jedoch der bereits erwähnte, für den biotechnologischen Einsatz interessante Aspekt der Steuerung durch Erhöhung der Expression der AT-Komponente sein, und zwar dergestalt, dass die AT-Aktivität die Engpass-Stelle („bottleneck“) der Biosynthese darstellt und somit auf ein Signal hin die Produktion des jeweiligen Sekundärmetaboliten durch Steigerung allein der Menge an AT-Aktivität schnell erhöht werden kann.

Die modularen PKS-I mit integrierter AT-Domäne bilden über den gemeinsamen Vorläufer, der durch den Verzweigungspunkt 9 repräsentiert wird, ein gemeinsames Cluster, welches daneben zwei als Schwestergruppen anzusehende Subcluster enthält, nämlich die PKS-I aus Pilzen und die FAS-I aus Tieren, welche über den gemeinsamen Vorläufer am Verzweigungspunkt 10 miteinander verbunden sind. Auf die Evolution der modularen PKS-I innerhalb der Bakterien wird im Abschnitt 4.2 der Diskussion ausführlich eingegangen. Zunächst soll hier der Zusammenhang zwischen diesem Typ von bakteriellem Enzymsystem und seinen eukaryotischen Pendants diskutiert werden.

4.1.2 Das System der eukaryotischen FAS und Zusammenhänge mit bakteriellen PKS des Typs I

Bedenkt man die große Bedeutung von Fettsäuren als Bestandteile von Zellmembrane, als Energiespeicherstoff und Vorläufer von Signalmolekülen und das dementsprechend ubiquitäre Auftreten dieser Gruppe von Primärmetaboliten in allen Organismenreichen, so mutet die strukturelle Heterogenität der zugrunde liegenden Synthasen in den taxonomischen Reichen der Eukaryoten geradezu erstaunlich an. In der folgenden Diskussion werden die evolutionären Positionen dieser Synthasen erörtert. Dabei liegt das Augenmerk insbesondere auf den beiden Fragen, ob es eine Übertragung von Genmaterial im Sinne eines horizontalen Gentransfers (HGT) von einem Bakterium auf einen frühen Eukaryoten gab und –falls dies zu bejahen ist – ob sich abschätzen lässt, wann dieser Transfer stattgefunden haben könnte. Der Abschnitt schließt mit der Darstellung einer umfassenden Theorie zur Evolution der FAS.

4.1.2.1 Varianten eukaryotischer FAS und ihre phylogenetischen Positionen

Eukaryotische Organismen besitzen strukturell ganz unterschiedliche Enzymsysteme zur Biosynthese von Fettsäuren (siehe hierzu auch **Abb. 4** im einleitenden Teil dieser Arbeit). Dies steht im Gegensatz zu anderen zentralen Stoffwechselwegen, etwa aus dem Kohlenhydrat-, Nucleotid- und Aminosäure-Stoffwechsel, deren Enzyme in den verschiedenen eukaryotischen Gruppen völlig identisch sind.

Die taxonomische Einteilung der Eukaryoten ist in den letzten Jahren durch neue Erkenntnisse über die evolutionären Zusammenhänge immer wieder Änderungen unterworfen gewesen und bislang nicht zufrieden stellend oder gar abschließend geklärt, wie sich aus der vielfältigen Literatur zu diesem Thema ergibt (siehe etwa die Darstellungen und Diskussionen in (Baldauf *et al.* 2000; Woese 2000; Cavalier-Smith 2002; Martin *et al.* 2003; Adl *et al.* 2005; Keeling *et al.* 2005; Embley und Martin 2006) und speziell zum Problem, die phylogenetische Wurzel des eukaryotischen Baums zu finden, die Diskussionen in (Dacks und Doolittle 2001; Stechmann und Cavalier-Smith 2002; Baptiste und Brochier 2004)). Zur Vereinfachung der Darstellung wird vorliegend hauptsächlich von der früher gebräuchlichen Einteilung in die Reiche Pflanzen, Tiere, Pilze und Protisten Gebrauch gemacht, da sie dem Zweck der Darstellung genügt. Dies geschieht in dem Bewusstsein, dass die sehr heterogene Gruppe der „Protisten“ keine taxonomisch sinnvolle Einteilung darstellt.

Bezogen auf diese „klassischen“ Reiche der Eukaryoten stellt sich die Situation folgendermaßen dar (zusammengefasst in **Tab. 14**):

Pflanzen synthetisieren ihre Fettsäuren in den Chloroplasten (Givan 1983), wobei hierfür das plastidäre, aus der Endosymbiose mit einem Cyanobakterium hervorgegangene Enzymsystem benutzt wird. Pflanzen ziehen also die typisch bakteriellen, distinkten Proteine der FAS vom Typ II zur Fettsäurebiosynthese heran. Die für diese Proteine codierenden Gene befinden sich zwar im Genom des Zellkerns, ihre Genprodukte werden jedoch posttranslational in die Plastiden verbracht (McFadden 1999). Die Gene wurden im Verlauf der Pflanzenevolution in den Zellkern transferiert (Martin *et al.* 1998). Ihre Abstammung von Cyanobakterien-ähnlichen Vorläufern ist durch viele Studien belegt (siehe z. B. (Martin *et al.* 2002)). Sie ergibt sich auch eindeutig aus den in dieser Arbeit vorgelegten Phylogenien, wie aus **Abb. 8** und aus **Abb. 11** hervorgeht. Außerdem finden sich in Pflanzen mitochondrial lokalisierte Enzyme der FAS-II. Die entsprechende KS konnte kürzlich in *Arabidopsis thaliana* identifiziert und anhand des rekombinanten Proteins biochemisch in Grundzügen charakterisiert werden (Yasuno, von Wettstein-Knowles und Wada 2004). Fettsäuresynthese im Cytosol findet in Pflanzen nicht statt.

Pilze produzieren die benötigten Fettsäuren hingegen im Cytosol mit Hilfe einer FAS des Typs I, also des fusionierten Multienzym-Typs. Dieses System ist mittlerweile durch die Aufklärung seiner Kristallstruktur (Jenni *et al.* 2006) sehr gut charakterisiert. Die beiden Untereinheiten sind durch die Domänenabfolgen AcT-ER-DH-AT für das Protein FAS1 (d.h. die β -Untereinheit) und ACP-KR-KS-PPT für das Protein FAS2 (d. h. die α -Untereinheit) gekennzeichnet, wobei AcT für eine spezielle Acetyl-Transferase und PPT für die Phosphopantetheinyl-Transferase steht. Auf die extrem auffälligen Übereinstimmungen zu den FAS-I aus bestimmten Bakterien (CMN-Gruppe) sowie die aus der phylogenetischen Rekonstruktion (**Abb. 8**) deutlich werdende evolutionäre Nähe wurde bereits in Abschnitt

4.1.1.2 eingegangen. Sie spielt für die Verhältnisse bei den Eukaryoten und damit für die Diskussion in diesem Abschnitt keine Rolle, da gemäß der weiter oben dargelegten Hypothese die Ursache dieser Übereinstimmungen in einem HGT-Ereignis in Richtung des prokaryotischen Vorläufers liegt. Betont werden muss hingegen die im Vergleich zu allen anderen fusionierten Multienzymen eigenständige phylogenetische Position der FAS-I aus Pilzen. Daneben verfügen Pilze ebenso wie Pflanzen zusätzlich über ein mitochondrial lokalisiertes FAS-II-System (siehe das rot unterlegte Cluster im phylogenetischen Baum der **Abb. 11**).

Tiere besitzen eine FAS-I mit der Domänenabfolge KS-AT-DH-ER-KR-ACP-TE. Diese Reihung entspricht exakt den Verhältnissen, die man bei den bakteriellen PKS-I findet. Wie im Fall der Pilze ist der FAS-Komplex im Cytosol lokalisiert. Auch Tiere verfügen über eine mitochondriale FAS-II, die auf das zur Entstehung dieses Organells hinführende Endosymbiose-Ereignis zurückzuführen ist. Erste Untersuchungen der KS-Komponente dieses Systems liegen vor (Zhang *et al.* 2005). Die cytosolische FAS-I der Tiere bildet eine phylogenetische Schwestergruppe mit den PKS-I aus Pilzen (siehe **Abb. 8** und **Abb. 11** im Ergebnisteil, Abschnitt 3.1.1, sowie das vereinfachte Schema in **Abb. 52**).

Die Situation bei den **Protisten** ist überaus vielgestaltig. Man findet die Sequenzen ihrer FAS-Komponenten über den gesamten phylogenetischen Baum verteilt (siehe **Abb. 11**), wodurch die Komplexität der dahinter stehenden evolutionären Verhältnisse klar zum Ausdruck kommt. Obwohl die nutzbare Datenbasis für dieses „Eukaryotenreich“ immer noch recht beschränkt ist, lassen sich eine Reihe von Variationen ausmachen, die für einen guten Teil der betrachteten Organismen vermutlich im Zusammenhang mit den bei parasitischer Lebensweise typischen Reduktionstendenzen zu sehen sind. Es fällt auf, dass in den meisten Fällen kein Nebeneinander von verschiedenen FAS-Typen vorkommt. Die Apicomplexa mit ihrem Vertreter *Plasmodium falciparum* und ebenso *Thalassiosira pseudonana* aus der Abteilung *Diatomea* scheinen ausschließlich über das quasi plastidäre FAS-II zu verfügen. Dieses resultiert aus der Aufnahme einer eukaryotischen Alge, deren Plastiden als so genannte „sekundäre Plastiden“ erhalten wurden (Cavalier-Smith 2002). In den Genomen der Vertreter der *Kinetoplastea*, Trypanosomen und Leishmanien, hingegen ließen sich nur FAS-II der mitochondrialen Linie identifizieren. Betrachtet man die Verhältnisse bei den Protisten im Gesamten, so ist augenfällig, dass nur zwei von ihnen über fusionierte Systeme mit aneinander gereihten funktionellen Domänen verfügen, nämlich *Cryptosporidium hominis*, das den Apicomplexa zugeordnet wird, und die Amöbe *Dictyostelium discoideum*. Cryptosporidien stellen insofern eine große Ausnahme im Vergleich zu den anderen Apicomplexa dar, als bekannt ist, dass eine sehr große Zahl an Genen mittels HGT aus Prokaryoten eingeschleust und in das Genom integriert wurden (Huang *et al.* 2004; Xu *et al.* 2004) und das Mitochondrium einer umfassenden Reduktion unterzogen wurde (Henriquez *et al.* 2005). In *Cryptosporidium parvum* wurde vor einiger Zeit ein PKS-ähnliches, sehr großes Protein mit mehr als 13.000 Aminosäure-Resten gefunden (Zhu *et al.* 2002). Die Sequenzen der darin enthaltenen KS-Domänen sind phylogenetisch in der Tat eng mit denjenigen aus bakteriellen PKS-I assoziiert, was sich am leichtesten durch einen solchen HGT erklären lässt. Die Konstellation bei Cryptosporidien stellt nach dem bisherigen Kenntnisstand nach jedenfalls eine Ausnahmesituation dar.

Wenngleich aufgrund der basalen Natur der Fettsäure-Biosynthese durchaus daran gedacht werden kann, die Evolution der FAS innerhalb der diversen Protisten-Gruppen heranzuziehen, um einen Beitrag zur Aufklärung der frühen Eukaryoten-Evolution zu leisten, sollen die hier kurz diskutierten

phylogenetischen Ergebnisse in der vorliegenden Arbeit nur dazu dienen, spezifische Schlussfolgerungen hinsichtlich der Evolution der FAS-I aus Tieren zu ermöglichen, wie weiter unten in Abschnitt 4.1.2.3 dargelegt wird.

Tab. 14: Übersicht zu den in Eukaryoten vorhandenen FAS- und PKS

Eukaryotische Gruppe	FAS	PKS-I
Pflanzen	FAS-II (Plastiden) FAS-II (Mitochondrien)	keine PKS-I
Pilze	FAS-I (Cytosol) FAS-II (Mitochondrien)	PKS-I (Cytosol)
Tiere	FAS-I (Cytosol) FAS-II (Mitochondrien)	keine PKS-I
Protisten	FAS-I oder FAS-II, diverse Lokalisierungen	In manchen Gruppen PKS-I-ähnliche Enzymsysteme

Von großer Bedeutung ist, dass sich die FAS-I aus Pilzen, wie bereits mehrfach erwähnt, in praktisch allen Belangen von der FAS-I der Tiere unterscheidet: Man findet eine völlig andere lineare Abfolge der Domänen auf der Polypeptid-Kette, es gibt keine Relationen zwischen den Längen der Interdomänen-Abschnitte, die Polypeptid-Ketten beherbergen teilweise enzymatische Domänen, die im jeweils anderen System nicht existieren, und der Aufbau der beiden Gesamtkomplexe zeigt keine Ähnlichkeiten. Während die FAS der Pilze nämlich als $\alpha_6\beta_6$ -Komplex organisiert ist (Jenni *et al.* 2006), bilden die Proteine der FAS-I aus Tieren ein Homodimer aus (Maier, Jenni und Ban 2006). Die vorgestellten phylogenetischen Rekonstruktionen ergaben eindeutig, dass hinsichtlich aller Domärentypen kein enger evolutionärer Zusammenhang zwischen diesen beiden FAS-Typen besteht (siehe u. a. **Abb. 9**).

4.1.2.2 *Stammt die FAS-I der Tiere von bakteriellen PKS-I ab?*

In der wissenschaftlichen Literatur wurde immer wieder auf die funktionellen und strukturellen Übereinstimmungen zwischen der FAS-I der Tiere und bakteriellen PKS-I hingewiesen. So hebt beispielsweise bereits ein früher Übersichtsartikel mit dem Titel „Molecular Genetics of Polyketides and its Comparison to Fatty Acid Biosynthesis“ (Hopwood und Sherman 1990) sehr stark auf diesen Umstand ab. Zuletzt gab es im Zusammenhang mit der Aufklärung der Kristallstrukturen der FAS aus Pilzen (Jenni *et al.* 2006) und Tieren (Maier, Jenni und Ban 2006) mehrere Kommentare (z. B. von (Smith 2006), wobei jedoch weder diese noch die Strukturartikel selber auf evolutionäre Fragestellungen eingehen. Die einzige Veröffentlichung, die sich dieses Themas am Rande annahm, liegt nunmehr etwa 15 Jahre zurück (Amy *et al.* 1992). Berichtet wurde darin über die erste vollständige DNA-Sequenz einer FAS aus Säugetieren, nämlich aus *Rattus norvegicus*, und deren Exon-Intron-Muster analysiert. In diesem Zusammenhang erwähnen die Autoren die auffallende Ähnlichkeit mit bakteriellen PKS (wobei PKS des Typs I gemeint

sind) und schlagen als mögliche Erklärung vor, dass die Genfusionsprozesse entweder bereits vor der „Trennung“ der prokaryotischen und eukaryotischen Linien stattgefunden haben oder es später zu einem HGT-Ereignis kam, bei dem das fusionierte FAS-I-System von einem Eukaryoten auf einen Prokaryoten übertragen wurde.

Dank der mittlerweile deutlich vergrößerten Datenbasis im Hinblick auf Sequenzen bakterieller PKS und infolge der zahlreichen neuen Erkenntnisse zur Evolution der Eukaryoten im Allgemeinen lassen sich die in Frage kommenden Möglichkeiten jetzt weitaus präziser fassen, wobei als weitere Enzymklasse inzwischen die PKS-I aus Pilzen hinzu gekommen sind. Die Hinweise auf einen gemeinsamen Ursprung der bakteriellen PKS-I auf der einen Seite und tierischen FAS-I und PKS-I aus Pilzen auf der anderen Seite, und zwar bereits in der Form eines fusionierten Systems, sind folgende: (1) Alle diese Systeme zeigen eine völlig übereinstimmende Abfolge ihrer Domänen; (2) die Länge der entsprechenden Interdomänen-Bereich ist in allen Fällen sehr ähnlich, und darüber hinaus zeigen diese Abschnitte auch auffallende Homologie auf Aminosäure-Ebene; (3) der Gesamtenzymkomplex ist nach vorliegenden Erkenntnissen auf gleiche Weise organisiert, nämlich als Homodimer, was für die FAS-I aufgrund der gelösten Kristallstruktur bekannt ist (Maier, Jenni und Ban 2006) und für die PKS-I zumindest als wahrscheinlichste Hypothese angenommen werden kann (siehe hierzu die Diskussion in (Staunton und Weissman 2001)); (4) die im Ergebnisteil dieser Arbeit vorgestellten phylogenetischen Rekonstruktionen (siehe Abschnitt 3.1.1) zeigen ganz eindeutig die Verknüpfung der einzelnen Cluster über einen nahen gemeinsamen Vorläufer auf. Zusammengenommen kann deshalb kein vernünftiger Zweifel daran bestehen, dass sich diese drei Multienzym-Systeme nicht unabhängig voneinander entwickelt haben, sondern vielmehr aus evolutionärer Sicht dasselbe darstellen, wie in der oben erwähnten Arbeit für bakterielle PKS-I und FAS-I aus Tieren bereits formuliert wurde (Amy *et al.* 1992).

Die dort angeführte alternative Erklärung durch einen HGT von einem Eukaryoten in Richtung einer prokaryotischen Zelle muss aus heutiger Sicht fallengelassen werden. Die Beobachtung, dass ein erheblicher Anteil der vollständig sequenzierten Bakteriengenome Gene für PKS-I birgt und diese „positiven“ Genome fast alle Bakteriengruppen, wenn auch in sehr unterschiedlicher Ausprägung, betreffen (siehe Abschnitt 3.2.1), spricht eindeutig gegen ein solches Szenario. Andererseits muss im Hinblick auf die dort vorgeschlagene Haupterklärung bedacht werden, dass FAS-I des tierischen Typs eben nicht ubiquitär in Eukaryoten anzutreffen sind (weshalb die Bezeichnung „eukaryotische FAS“ definitorisch keinen Sinn macht). Insbesondere kommt sie in Pflanzen überhaupt nicht vor und im Bereich der Protisten nur äußerst vereinzelt. Deswegen erscheint das Auftreten dieser Systeme weit vor der Entstehung der Eukaryoten nicht als einfache Erklärung. Sie wäre nur über die Annahme einer sehr großen Zahl von Fällen, in denen diese Art von FAS wieder eliminiert wurde, aufrecht zu erhalten.

Bei einem so tiefgreifenden möglichen Transferereignis wie im Fall der FAS-I könnte man auch an eine Kopplung mit der Aufnahme eines Endosymbionten denken, aus dem sich das mitochondriale Organell entwickelte. Denn diese Aquisition ist bekanntermaßen ebenso wie die spätere „Vereinnahmung“ eines Cyanobakteriums unter Ausbildung der pflanzlichen Plastiden der Hauptgrund dafür, weshalb in den Genomen von Eukaryoten quasi-bakterielle Gene zu finden sind. Allerdings stößt man bei dieser Annahme auf ähnliche Schwierigkeiten, wie sie eben erläutert wurden. Da Mitochondrien nach herrschender Meinung im Verlauf der Zellevolution nur ein einziges Mal entstanden sind (Gray, Burger und Lang 1999; Lang, Gray und Burger 1999), würden in diesem Szenario zunächst alle Eukaryoten zusammen mit dem

Endosymbionten die bakterielle PKS-I erhalten haben. Man mag zwar deren Fehlen in Pflanzen mit den offensichtlichen Veränderungen im Stoffwechsel der Fettsäuren nach Ausbildung der Plastiden in etwas unbestimmter Art und Weise erklären können, die Situation im Bereich der diversen Protisten hingegen nicht. Außerdem lässt sich aus den Bäumen in **Abb. 8**, **Abb. 9** und **Abb. 10** nicht erkennen, dass PKS-I aus α -Proteobakterien den FAS-I aus Tieren und den PKS-I aus Pilzen phylogenetisch näher lägen als PKS-I aus anderen Bakteriengruppen.

Im Allgemeinen sollte derjenigen Erklärung der Vorzug gegeben werden, die die geringste Zahl an zusätzlichen Annahmen erfordert (Parsimonie-Prinzip). Im vorliegenden Fall bedeutet das: Am wahrscheinlichsten ist, dass es nach der Etablierung der eukaryotischen Entwicklungslinie zu einem Gentransfer der PKS-I-codierenden Sequenzen von einem Bakterium auf einen Vorläufer von Tieren und Pilzen kam. In der Linie der Tiere, zu der einige Protisten wie die *Dictyosteliida* in enger Verwandtschaft stehen (Keeling *et al.* 2005), entwickelte sich aus dem aus einem Prokaryoten übertragenen System die FAS, während sie in Pilzen eine – iterativ arbeitende – PKS blieb. Die Einfachheit dieser Hypothese lässt sich anhand aktueller Eukaryoten-Phylogenien belegen. Die **Abb. 53** zeigt eine zusammenfassende Skizze der evolutionären Beziehungen zwischen den eukaryotischen Hauptzweigen, die den derzeitigen Kenntnisstand wiedergibt (Embley und Martin 2006). Aus dieser Darstellung geht hervor, dass der gemeinsame Vorläufer der Opisthokonten ein guter Kandidat für den vorgeschlagenen Gentransfer ist. Die Gruppe der Opisthokonten umfasst neben den Tieren und Pilzen auch diverse Amöben, zu den *D. discoideum* gehört. Die enge evolutionäre Verknüpfung zwischen Tieren und Pilzen, die sie zugleich deutlich von anderen eukaryotischen Linien abhebt, ist eine einzigartige 12 Aminosäure-Reste umfassende Insertion in der Sequenz des Elongationsfaktors EF-1 α (Baldauf und Palmer 1993). Die enge phylogenetische Beziehung der Tier/Pilz-Gruppe zu *D. discoideum* als Mitglied einer direkten Schwestergruppe wiederum konnte kürzlich nachgewiesen werden (Steenkamp, Wright und Baldauf 2006). Im Gegensatz zu *D. discoideum*, das nach geltender Taxonomie in das Subphylum *Mycetozoa* eingegliedert ist, gehört *Entamoeba histolytica* aus dem Subphylum *Sarcodina* nicht in diese Schwestergruppe, sondern zu den *Amoebozoa* im engeren Sinne, die sich in **Abb. 53** aus dem Verzweigungspunkt 2 als eigene Linie von den Opisthokonten abspalten. Dies würde erklären, warum *E. histolytica* keine FAS-I besitzt.

Die hier vorgeschlagene Hypothese wird in Zukunft mit der steigenden Verfügbarkeit von Genomsequenzen aus Protisten weiter überprüft werden können. Entscheidend wird dabei sein, welche Verteilungsmuster von FAS-I und FAS-II in den verschiedenen Gruppen zu Tage treten und wie sich die Gruppe selbst in phylogenetischer und taxonomischer Hinsicht stabilisieren lassen.

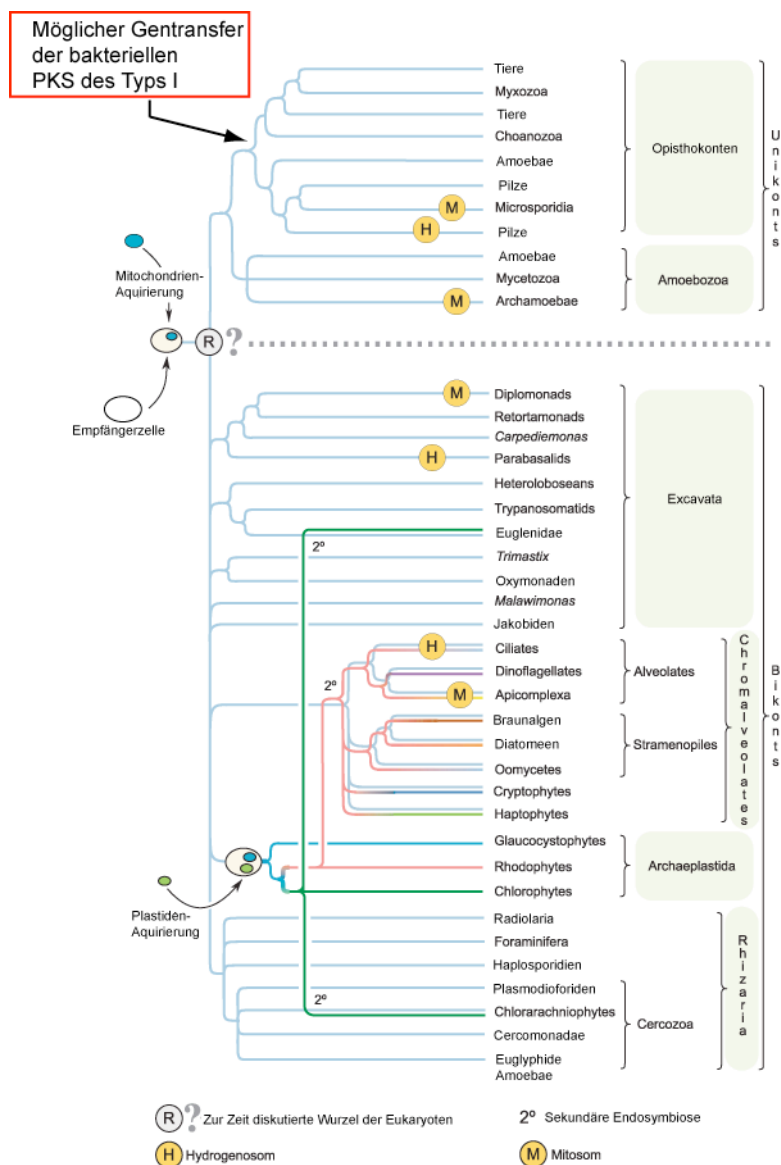


Abb. 53: Zusammenfassende Skizze aktueller Hypothesen zu den phylogenetischen Beziehungen zwischen den eukaryotischen Hauptgruppen.

Der Baum ist eine Zusammenfassung der Ergebnisse verschiedener Arbeitsgruppen, die kürzlich in einem umfassenden Übersichtsartikel zusammengefasst wurden (Adl *et al.* 2005). Der mögliche Vorläufer, in den eine bakterielle PKS-I mittels HGT hinein gelangte, ist mit einem Pfeil kenntlich gemacht. Mit H und M sind Zweige gekennzeichnet, die Organismen mit gut untersuchten Hydrogenosomen bzw. Mitosomen enthalten. Mit R ist die derzeit diskutierte Lage der möglichen Wurzel der Eukaryoten gekennzeichnet. (Abbildung aus Embley und Martin 2006, verändert)

4.1.3 Evolutionäre Beziehungen zwischen den PKS-III aus unterschiedlichen Organismengruppen

Bei den PKS des Typs III kommt es im Gegensatz zu den PKS-I und PKS-II nicht so sehr auf eine Einordnung in die Gesamtevolution von PKS und FAS an, da die Sachlage weit weniger kompliziert ist. Die vergleichsweise einfach gebauten PKS-III leiten sich, wie sich aus Sequenzvergleichen ergibt, eindeutig von der FabH-Komponente der FAS-II ab (zusammengefasst in (Austin und Noel 2003)). Bei einer auf die Evolution ausgerichteten Betrachtung der PKS-III kommt es stattdessen auf die Frage an, ob Aussagen darüber gemacht werden können, in welcher Organismengruppe diese Enzyme zuerst auftraten. Dies ergibt sich aus dem Umstand, dass man lange Zeit annahm, PKS-III kämen ausschließlich in höheren Pflanzen vor, wären also gewissermaßen eine ganz spezifisch pflanzliche „Eigenentwicklung“. Diese Auffassung änderte sich ab Mitte der 90er Jahre des 20. Jahrhunderts mit der Entdeckung eines PKS-III-ähnlichen Enzyms in dem Actinobakterium *Streptomyces griseus* (Ueda *et al.* 1995). Dieses Enzym katalysiert, ausgehend von fünf Malonyl-CoA-Einheiten, die Bildung von 1,3,6,8-Tetrahydronaphthalen, welches spontan zu Flaviolin oxidiert (Ueda *et al.* 1995). In den folgenden Jahren wurden zahlreiche weitere bakterielle PKS-III identifiziert, außerdem auch einige Beispiele in Pilzen (Bangera und Thomashow 1999; Funa *et al.* 1999; Moore und Hopke 2001; Izumikawa *et al.* 2003; Saxena *et al.* 2003).

Die in Abschnitt 3.1.2 vorgestellte Analyse der phylogenetischen Verteilung von FabH- und PKS-III-Sequenzen aus unterschiedlichen Organismengruppen (siehe **Abb. 12**) lässt darauf schließen, dass die PKS vom Typ III zuerst in Bakterien aus dem FabH-Protein entstanden und dass diese prokaryotischen PKS-III die Vorläufer aller heute in Pflanzen zu findenden PKS-III-Typen sind. Insofern wäre der Baum in **Abb. 12** als Ausdruck einer kontinuierlichen Evolution „von unten nach oben“ zu interpretieren. Wollte man hingegen am traditionellen Bild der PKS-III als originär pflanzliches Enzymsystem festhalten und damit auch implizieren, dass sie in Pflanzen entstanden sind, so müssten eine Reihe von horizontalen Gentransfers (HGT) herangezogen werden, um ihr Vorkommen in Pilzen und Bakterien zu erklären. Diese Alternativerklärung mag noch statthaft gewesen sein, als erst sehr wenige nicht-pflanzliche Sequenzen bekannt waren, so dass durchaus von vereinzelten Gentransfers ausgegangen werden konnte. Mit der im Zuge der vielen Sequenzierungsprojekten zu bakteriellen Genomen stetig steigenden Zahl neuer PKS-III-Sequenzen, die nicht aus Pflanzen stammen, wird diese Erklärung jedoch immer unwahrscheinlicher. Dies ergibt sich auch in mehrfacher Hinsicht aus den phylogenetischen Zusammenhängen der **Abb. 12**:

- (1) Es fällt auf, dass die PKS-III aus Bakterien eine deutlich größere Nähe zu den Vorläuferenzymen FabH sowohl aus Bakterien als auch aus Pflanzen zeigen. Die Aminosäure-Sequenzen der pflanzlichen PKS-III liegen dagegen phylogenetisch in deutlicher Entfernung.
- (2) Weiterhin ist zu beachten, dass bakterielle PKS-III eine viel größere Diversität als ihre pflanzlichen Pendanten aufweisen und nur eine einzige Untergruppe der prokaryotischen Sequenzen relativ nahe am Pflanzen-Cluster liegt (Gruppe III der Bakterien). Interessanterweise enthält diese Untergruppe auch bislang uncharakterisierte Sequenzen aus Genomen von Cyanobakterien. Dies mag die Möglichkeit verdeutlichen, dass die heute quasi ubiquitär in Pflanzen zu findenden PKS-III aus Genmaterial des Cyanobakterien-ähnlichen Endosymbionten hervorgegangen sein könnten, der sich zum Chloroplasten entwickelte.

- (3) Schließlich muss die Verteilung der PKS-III innerhalb der Bakterien bedacht werden. Sie werden mittlerweile in allen Hauptgruppen der Bakterien gefunden, wenn auch jeweils eher vereinzelt als generell. Nur eine kleine Minderheit der prokaryotischen Sequenzen des zugrunde liegenden Datensatzes stammt aus Pflanzen-assoziierten Bakterien.

In der Gesamtschau gibt es somit eine Reihe von guten Gründen, aufgrund derer die Entstehung der PKS-III in Prokaryoten und der spätere Transfer zu den Pflanzen gegenüber der umgekehrten Konstellation zu favorisieren wäre. Allerdings kann die Einschätzung der evolutionären Entwicklungslinien zum jetzigen Zeitpunkt nur einen vorläufigen Charakter haben, da die Mehrzahl der in der Untersuchung verwendeten bakteriellen Sequenzen sich auf Proteine bezieht, die bislang gänzlich uncharakterisiert sind, darunter gerade auch diejenigen mit der engsten Verwandtschaft zu den pflanzlichen Enzymen.

4.2 Evolution der modularen PKS-I innerhalb der Bakterien

In diesem Abschnitt wird anfangs die in **Abb. 51** als Stufe 2 bezeichnete Untersuchungsebene diskutiert und darauf aufbauend – gewissmaßen als noch besser aufgelöster „Vergrößerungsschritt“ – zur Stufe 3 vorangeschritten. Der Gliederung im Ergebnisteil entsprechend richtet sich die Betrachtung zunächst auf die Verteilung der PKS vom Typ I auf die unterschiedlichen Bakteriengruppen, verbunden mit der Frage, ob sich bestimmte Bakterien-spezifische Kriterien finden lassen, die mit der Verteilung korrelieren (Abschnitt 4.2.1). Danach folgt die Diskussion der in den phylogenetischen Rekonstruktionen der KS- und AT-Domänen gefundenen Muster und der möglichen Schlussfolgerungen hinsichtlich erkennbarer typischer Evolutionsabläufe (Abschnitt 4.2.2). Schließlich beschäftigt sich der Abschnitt 4.2.3 mit den Fragen der Mechanismen der PKS-I-Evolution innerhalb eines Organismus, womit, folgt man dem Schema in **Abb. 51**, die Stufe 3 der möglichen Untersuchungsebenen erreicht ist. Hier geht es um intragenomische Aspekte der PKS-Entwicklung und insbesondere die Frage, auf welche Weise ein Organismus das besondere modulare Bauprinzip zur Generierung von metabolischer Strukturdiversität ausnutzen kann.

4.2.1 Verteilung der PKS-I in Bakterien

Für PKS-I-Proteine codierende Sequenzen konnten ausschließlich in eubakteriellen Genomen identifiziert werden, die Suche in den zur Verfügung stehenden Genomen von Archaea brachte hingegen keine positiven Resultate (siehe Abschnitt 3.2.1). Eine Erklärung dafür könnte in einer Besonderheit der archaebakteriellen Fettsäure-Biosynthese zu suchen sein. Bislang konnte nämlich in keinem Archaea-Genom ein zu FabD homologes Protein aufgefunden werden, obwohl alle anderen Proteine bzw. Enzymaktivitäten einer bakteriellen FAS-II vorhanden sind (Pereto, Lopez-Garcia und Moreira 2004). Es wird vermutet, dass die entsprechende AT-Aktivität in Form von FabD in der Linie der Archaea zu einem frühen Zeitpunkt verloren ging und durch ein nicht-homologes Enzym ersetzt wurde (Pereto, Lopez-Garcia

und Moreira 2004). Insofern konnten die Archaea möglicherweise keine eigenen PKS-Systeme entwickeln, da ihnen die AT-Komponente fehlte, die notwendig gewesen wäre, um sie zu „konstruieren“.

Etwa 20 % der verfügbaren Genome enthielten mindestens ein Gen einer vollständigen PKS-I, d. h. die zusammenhängende Anordnung KS-AT-ACP. Dieser Wert bezieht sich auf eine bakterienweite Betrachtung ohne Differenzierung nach einzelnen Gruppen. Tatsächlich zeigen die in **Tab. 5** zusammengetragenen Werte jedoch, dass die Anteile der „positiven“ Genome je nach Bakteriengruppe sehr unterschiedlich ausfallen. An dieser Stelle sei nochmals angemerkt, dass klare statistische Aussagen hierüber nicht möglich sind, da die Zahl an verfügbaren Genomdaten im Hinblick auf die verschiedenen Bakteriengruppen sehr unterschiedlich ausfiel. So konnten 60 proteobakterielle Genome und 39 von Firmicuten herangezogen werden, während nur zwei komplette Streptomyceten-Genome und nur vier vollständige Genome aus der Gruppe der Spirochaeten für die Auswertung zur Verfügung standen. Außerdem ist zu berücksichtigen, dass in manchen Gruppen eine Reihe von sehr eng verwandten Bakterien sequenziert wurden, häufig deshalb, weil es sich um medizinisch relevante humanpathogene Spezies und Stämme handelt wie im Fall der Mycobakterien, in deren Gruppe z. B. zwei nur in wenigen Details voneinander abweichende Stämme derselben Spezies Teil des Datensatzes waren. Daher handelt es sich im Folgenden um eine eher qualitativ geprägte Betrachtung.

Grundsätzlich ließ sich eine – auch statistisch signifikante – Korrelation zwischen Genomgröße und Zahl an PKS-Genen feststellen (dargestellt in **Abb. 13**). Dahinter verbirgt sich augenscheinlich ein Zusammenhang zwischen der Zahl an PKS-Genen und der „Komplexität“ eines Organismus, gerade auch unter Berücksichtigung seiner Lebensweise und seines ökologischen Umfelds und der dadurch bedingten Notwendigkeit, um Nährstoffe zu konkurrieren, Signale zu detektieren, mit anderen Organismen zu kommunizieren usw. Beispielsweise findet man in der Gruppe der Cyanobakterien mit 22 codierten PKS-I-Proteinen an führender Stelle *Nostoc punctiforme*. Es verfügt über ein im Vergleich mit anderen Cyanobakterien sehr großes Genom mit einer Länge von mehr als neun Mbp und weist neben der Möglichkeit, verschiedene Zelltypen auszudifferenzieren auch eine ausgeprägte Fähigkeit zur Ausbildung symbiontischer Beziehungen auf (siehe hierzu die Einleitung, Abschnitt 1.6.2). *N. punctiforme* ist ein mehrzelliges Cyanobakterium mit terrestrischer Lebensweise, das aufgrund dessen mit stark variierenden Umweltbedingungen konfrontiert ist. Damit unterscheidet es sich hinsichtlich seiner „Komplexität“ deutlich von den anderen im Datensatz enthaltenen Cyanobakterien wie etwa den *Prochlorococcus*-Stämmen, die keine PKS-I besitzen, oder *Gloeobacter violaceus*, dessen Genom für nur drei PKS-Proteine codiert. *Prochlorococcus* ist ein marin lebendes Bakterium, das sich auf bestimmte ökologische Nischen spezialisiert hat und über weitgehend konstante Lebensbedingungen verfügt (Partensky, Hess und Vaultot 1999).

Auffällig sind die hohen Anteile bei den Actinobakterien (10 positive von insgesamt 13 Genomen) und den Cyanobakterien (5 von 11). Zur Gruppe der Actinobakterien zählen auch die Mycobakterien, die bezüglich ihres Gehalts an PKS-Genen allerdings einen Sonderfall darstellen. Dies ist bereits daran erkennbar, dass die Mycobakterien eine deutliche Abweichung in der Korrelation von Genomgröße und Zahl an PKS-Genen aufweisen (siehe **Abb. 13**). Sie enthalten deutlich mehr PKS-Gene, als die Größe ihres Genoms erwarten lässt. Der Grund hierfür liegt wahrscheinlich in ihrer spezialisierten, parasitischen Lebensweise, die zwar in vielen Belangen zu einer metabolischen Reduktion geführt hat (Vissa und Brennan 2001), jedoch nicht in Bezug auf den Lipid-Stoffwechsel. In diesem Zusammenhang findet man

vielmehr eine besonders hohe enzymatische Diversität aufgrund der besonders strukturierten mycobakteriellen Zellwand, an deren Synthese einige der hier als PKS klassifizierten Enzyme beteiligt sind (Brennan und Nikaido 1995). Dies spiegelt sich auch in der extrem großen Zahl an Genen wider, die am Lipidstoffwechsel der Mycobakterien Anteil haben. *Mycobacterium tuberculosis* besitzt etwa 250 solche Gene, während in *Escherichia coli*, dessen Genom eine ähnliche Größe aufweist, nur ca. 50 Gene in diese Gruppe fallen (Cole *et al.* 1998).

Aufschlussreich kann die nähere Betrachtung solcher Konstellationen sein, bei denen in einer großen Bakteriengruppe nur wenige Vertreter über PKS-Enzymsysteme verfügen. Beispielsweise findet man in der Klasse der γ -Proteobakterien nur fünf Genome mit PKS-I-Genen, während 28 Genome keine solchen Gene besitzen. Die Familien *Pasteurellaceae*, *Vibrionaceae* und *Xanthomonadaceae* sind völlig frei von PKS des Typs I; sie treten nur innerhalb der *Enterobacteriaceae* und der *Pseudomonadaceae* auf. In der Familie der *Enterobacteriaceae* handelt es sich zum einen um das Genom des *Escherichia coli*-Stamms CFT073. Vier andere Stämme von *E. coli* sind hingegen negativ. Über den produzierten Metaboliten und seine physiologische Funktion ist nichts bekannt. Zum anderen existieren PKS-I-Gene in *Photobacterium luminescens*. Dieses Bakterium produziert mehrere gegen Insekten gerichtete Toxine und lebt in Symbiose mit einem Nematoden (Duchaud *et al.* 2003). Auch in diesem Fall ist über den entsprechenden Metaboliten noch nichts bekannt. Schließlich wurden PKS-I-Komponenten in *Yersinia pestis* gefunden, wo sie Teil des Enzymsystems zur Biosynthese des Siderophors Yersiniabactin sind (Miller *et al.* 2002). In der Familie der *Pseudomonadaceae* konnten PKS-I-Gene nur in *Pseudomonas syringae* identifiziert werden, wo sie an der Biosynthese des Pflanzentoxins Coronatin beteiligt sind, während sie in den ebenfalls vollständig sequenzierten Arten *Pseudomonas aeruginosa* und *Pseudomonas putida* nicht auftreten. Alle drei sind bodenlebende Bakterien, *P. syringae* verhält sich pflanzenpathogen, *P. aeruginosa* ist als opportunistischer Krankheitserreger beim Menschen bekannt. Klare Regelmäßigkeiten in der Verteilung der PKS-I-Gene lassen sich somit in der Klasse der γ -Proteobakterien nicht erkennen. Gemeinsam ist ihnen allen, dass sie vermutlich über horizontalen Gentransfer (HGT) eingebracht wurden (siehe dazu weiter unten die Diskussion in Abschnitt 4.2.2). Das Auffinden von Korrelationen zwischen Gehalt an PKS des Typs I und Eigenschaften der betreffenden Bakterienarten wird entscheidend dadurch erschwert, dass nur in den wenigsten Fällen Daten zur Funktion der jeweiligen Metabolite vorliegen.

Auch in der Abteilung *Firmicutes* finden sich PKS-I in nur sehr geringer Zahl, und zwar ausschließlich in der Ordnung *Bacillales*. Allgemeine, klar zu umreisende Verteilungsmuster sind jedoch in dieser Gruppe ebensowenig zu erkennen wie bei den γ -Proteobakterien. So weist das Genom von *Bacillus subtilis* (subsp. *subtilis*, Stamm 168) eine Reihe von PKS-I-Genen auf, wohingegen die eng verwandte Art *Bacillus cereus* über keine Gene dieses Typs verfügt, wobei zu betonen ist, dass es sich in beiden Fällen um weit verbreitete Bodenbakterien mit sehr ähnlicher Lebensweise handelt.

Der weiter oben in diesem Abschnitt bereits erwähnte grundsätzliche Zusammenhang zwischen „Komplexität“ des bakteriellen Organismus und der Zahl an PKS-I-Genen lässt sich bestätigen, wenn man die Betrachtung auf Bakterien ausdehnt, für die zum Zeitpunkt der hier vorgelegten Analyse noch keine vollständigen Genomsequenzen erhältlich waren. Ein besonders klares Beispiel für diesen Zusammenhang bieten die Myxobakterien. Sie werden taxonomisch innerhalb der Klasse der δ -Proteobakterien in die Ordnung *Myxococcales* eingruppiert. Myxobakterien zeigen eine Art von

„Sozialverhalten“, indem sie in Abhängigkeit von den Umweltbedingungen Fruchtkörper ausbilden, ein Vorgang, der durch chemische Kommunikation zwischen den Zellen koordiniert sein muss (Reichenbach 1999). Gleichzeitig sind Myxobakterien schon seit längerem dafür bekannt, dass sie eine reiche Quelle an Sekundärmetaboliten darstellen, von denen viele unter Beteiligung von PKS synthetisiert werden (Bode und Müller 2006). Dies wurde kürzlich auf der genetischen Ebene durch die Veröffentlichung des Genoms des Myxobakteriums *Myxococcus xanthus* auf eindrucksvolle Weise bestätigt (Goldman *et al.* 2006). Die Gencluster für PKS-Systeme und gemischte PKS/NRPS-Systeme machen 8,6 % des gesamten Genoms von 9,14 Mbp aus, umfassen also ca. 800 kbp. Damit ist die im Genom von *M. xanthus* codierte Kapazität zur Biosynthese der entsprechenden Verbindungen etwa doppelt so groß wie in den über ein Genom ähnlicher Größe verfügenden Streptomyceten *S. avermitilis* und *S. coelicolor* (Goldman *et al.* 2006).

Insgesamt lässt sich somit konstatieren: Die Interpretation der Verteilung von PKS-I-Genen in Bakterien ist aufgrund der häufig noch lückenhaften Informationen zu den entsprechenden Metaboliten und ihren biologischen Funktionen schwierig. Dennoch fällt auf, dass tendenziell diejenigen Bakterien, die eine relativ große Zahl dieser Gene besitzen, komplexe ökologische und morphologische Eigenschaften aufweisen wie etwa symbiotische Beziehungen zu anderen Organismen und Zelldifferenzierungsprozesse, wie sie von Actinobakterien, Myxobakterien und Heterocysten-bildenden Cyanobakterien bekannt sind (Reichenbach 1999; Meeks *et al.* 2002; Gehring *et al.* 2004).

4.2.2 Muster in der Evolution von PKS des Typs I in Bakterien

Die Untersuchung modularer PKS-I in Bakterien erbrachte folgende Hauptergebnisse:

- (1) Die beiden Varianten von AT-Domänen, die Malonyl-CoA bzw. Methylmalonyl-CoA als ihre Substrate benutzen, bilden zwei evolutionär klar getrennte Linien. In anderen Worten: Diese beiden Varianten sind in ihrem Ursprung eindeutig monophyletisch.
- (2) In den Gruppen der Actinobakterien, Cyanobakterien und Myxobakterien finden sich große Gencluster, deren KS- und AT-Domänen in den jeweiligen phylogenetischen Bäumen überwiegend in einem gemeinsamen Zweig liegen und am besten durch eine Reihe von aufeinander folgenden Duplikationsereignissen zu erklären sind. Hingegen finden sich innerhalb des Phylums *Proteobacteria* in den Klassen der α -, β - und γ -Proteobakterien klare Hinweise auf die Aquirierung dieser Gene mittels horizontalem Gentransfer (HGT).

Auf dieser Basis lässt sich ein Modell der Evolution der bakteriellen PKS des Typs I entwickeln, das im folgenden Unterabschnitt vorgestellt wird. Im Anschluss wird auf die besondere Bedeutung von HGT-Ereignissen und damit verbundene spezifische genomische Organisationsprinzipien, die so genannten Genomischen Inseln, eingegangen und als Spezialfall die Situation bei den PKS-I-Clustern innerhalb der Familie *Streptomycetaceae* beleuchtet.

4.2.2.1 Modell der Evolution von PKS-I in Bakterien

Das Modell der Evolution der bakteriellen PKS-I ist in **Abb. 54** anhand der AT-Domänen graphisch skizziert. Ausgangspunkt ist das jeweilige Enzym der bakteriellen Fettsäuresynthese, also im Fall der AT-Domänen das Protein FabD. Dessen Gen wurde zu einem sehr frühen Zeitpunkt dupliziert, und das paraloge Protein erfuhr eine funktionelle Spezialisierung, wodurch es die AT-Aktivität einer frühen PKS bereitstellte.

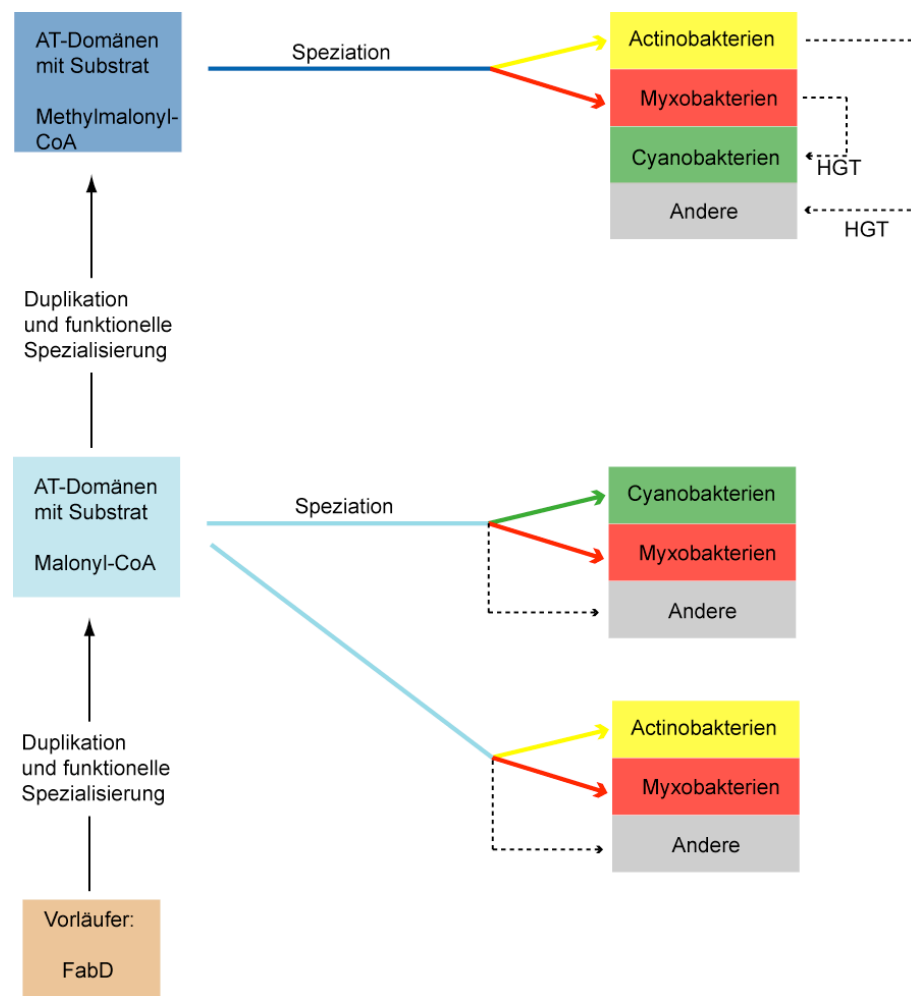


Abb. 54: Szenario der Evolution bakterieller PKS-I.

Dargestellt ist in der vertikalen Dimension von unten nach oben die stufenweise Evolution mittels Genduplikationen und funktioneller Spezialisierung der entstandenen Paraloge und in der horizontalen Dimension die Ausbildung von Orthologen mittels Speziation. Die gestrichelten Linien verweisen auf möglichen horizontalen Gentransfer (HGT). Weitere Erläuterungen werden im Text gegeben.

Im Zuge der Genfusionen, die zur Ausbildung des gemeinsamen Vorläufers der PKS des Typs I führten, wandelte sich das Paraloge schließlich zur AT-Domäne. Diese besaß in der ursprünglichen Form mit allergrößter Wahrscheinlichkeit die von Fettsäuresynthasen bekannte Spezifität für Malonyl-CoA, wie sich in

Abb. 14 aus der direkten Abfolge dieses AT-Typs nach den FabD-Proteinen aus diversen Bakteriengruppen ergibt. Da die Proteinsequenzen aus diesen Bakteriengruppen als Außengruppen der Phylogenie fungieren, ist der Baum der AT-Domänen in dieser Darstellung von unten nach oben gerichtet. Im Verlauf sich daran anschließender Speziationsprozesse bildeten sich die überwiegend nach bakteriellen Gruppen angeordneten Cluster von AT-Domänen heraus, in denen sich die Sequenzen aus Cyanobakterien, Actinobakterien und Myxobakterien in den meisten Fällen gut gegeneinander abgrenzen lassen. Diese Anordnungen decken sich mit den allgemein anerkannten monophyletischen Ursprüngen dieser Gruppen, wie sie sich auch aus dem 16S-DNA-basierten Baum der **Abb. 16** ergeben. Die allgemeine Tendenz, dass die Phylogenie der verschiedenen Domänentypen in Beziehung zu den Herkunftsorganismen steht und im Allgemeinen den Spezies-Phylogenien entspricht, wurde auch in einer anderen Studie unter Heranziehung der KS-Domänen nachgewiesen (Moffitt und Neilan 2003), allerdings mit einem deutlich kleineren Datensatz.

Eine weitere Duplikation mit nachfolgender funktioneller Spezialisierung führte zur Entstehung der anderen Hauptgruppe von AT-Domänen, die anstatt Malonyl-CoA das Substrat Methylmalonyl-CoA verwerten. Auch diese Ereignisse müssen zu einem sehr frühen Zeitpunkt der bakteriellen Evolution stattgefunden haben, da auffälligerweise eine völlig klare phylogenetische Trennung zwischen diesen beiden Typen von AT-Domänen besteht. Nach den derzeit vorliegenden Sequenzdaten spricht vieles dafür, dass die Methylmalonyl-CoA-benutzende Variante nur einmal in der Evolution entstanden ist. Diese klare Aufspaltung ergab sich auch bei Analysen zur Vorhersage der Substratspezifität von AT-Domänen (Yadav, Gokhale und Mohanty 2003). In dieser Untersuchung bestand der Datensatz im Vergleich zur hier vorgelegten Analyse jedoch fast ausschließlich aus PKS-Sequenzen von Streptomyceten, weshalb die Aussagekraft der dort vorgestellten Ergebnisse von weniger allgemeinem Charakter war. Weiterhin zeigte sich auch im Fall der Methylmalonyl-CoA-benutzenden AT-Domänen eine ausgeprägte Cluster-Bildung von Sequenzen derselben Bakteriengruppe, jedenfalls im Vergleich der Domänen aus Actinobakterien und Myxobakterien.

4.2.2.2 Einfluss von horizontalem Gentransfer auf die Evolution bakterieller PKS-I

Abgesehen von Duplikations- und Speziationsprozessen berücksichtigt das in **Abb. 54** skizzierte Modell auch die Beteiligung von HGT-Ereignissen, die durch gestrichelte Linien dargestellt sind. Die Situation lässt sich zwar nicht dahin gehend völlig verallgemeinern, dass alle Proteobakterien ihre PKS-Gene quasi sekundär mittels HGT erhalten hätten. Prinzipiell besteht immer auch die Möglichkeit eines originären proteobakteriellen Ursprungs, der durch die Beschränkungen im Datensatz verschleiert wird. Jedoch brachte die Auswertung der phylogenetischen Bäume der KS- und AT-Domänen ein klares Ergebnis hinsichtlich der Zahl an potenziellen HGT-Ereignissen (siehe **Tab. 7**). Darüber hinaus gibt es auch andere, von der phylogenetischen Rekonstruktion unabhängige Indizien für den Austausch genetischen Materials. Die Gene für PKS des Typs I in den Klassen der α -, β - und γ -Proteobakterien zeigen eine überaus unregelmäßige Verteilung (siehe **Tab. 5** und die Diskussion in Abschnitt 4.2.1). Im Falle der PKS-Gene aus *Pseudomonas syringae* und *Yersinia pestis* wurden konkret signifikante Abweichungen im GC-Gehalt gefunden, wenn man sie mit den für diese Bakterienklasse typischen genomweiten Werten vergleicht. Außerdem weist ein für α -, β - und γ -Proteobakterien charakteristisches Merkmal eindeutig darauf hin, dass

HGT bei der Herausbildung von PKS-Systemen in diesen Klassen eine wichtige Rolle gespielt hat: Die Gene liegen nämlich ganz überwiegend auf so genannten Genomischen Inseln (englisch: *genomic islands*) oder auf Plasmiden, sind also Teil mobiler genetischer Systeme, die den Austausch zwischen verschiedenen Organismen erleichtern. Das Konzept der Genomischen Insel hat seinen Ursprung in der Entdeckung der Pathogenitätsinseln in den Genomen prokaryotischer Krankheitserreger. Pathogenitätsinseln sind große genomische Bereiche, die folgende Eigenschaften aufweisen: (1) Sie codieren für ein oder mehrere Virulenz-assoziierte Genprodukte, (2) weisen einen GC-Gehalt auf, der von anderen Bereichen des Chromosoms abweicht, (3) sind häufig mit den Genen von tRNAs gekoppelt, (4) in den meisten Fällen von Sequenzwiederholungen in Form von Insertionssequenzen flankiert und kommen in pathogenen Stämmen im Vergleich zu eng verwandten nicht-pathogenen Varianten deutlich gehäuft vor (Dobrindt *et al.* 2004). Zu den Virulenz-assoziierten Genen gehören z. B. solche, welche die Biosynthese von Siderophoren ermöglichen und damit zu einer effizienten Aufnahme von Fe^{3+} -Ionen beitragen. Hierzu gehört auch das von *Y. pestis* produzierte Yersiniabactin. Später fand mehr vermehrt genomische Bereiche in nicht-pathogenen Bakterien, die – abgesehen von der Eigenschaft, Virulenzfaktoren zu codieren – allen Kriterien genügten. Offensichtlich können die durch diese Bereiche vermittelten physiologischen Funktionen auch in anderen Zusammenhängen wie etwa Symbiose, Resistenz gegen bestimmte Substanzen und Abbau von Xenobiotika wertvoll sein und die Fitness des Organismus erhöhen (siehe hierzu z. B. die Auflistung in (Dobrindt *et al.* 2004)). Deshalb wurde als Obergriff der Terminus Genomische Inseln eingeführt.

Die evolutionäre Bedeutung von Genomischen Inseln liegt auf der Hand. Im Gegensatz zum Transfer einzelner Gene werden mit Genomischen Inseln ganze „Gen-Pakete“ weitergegeben, was eine schnelle und erfolgreiche Nutzung durch den Rezipienten wahrscheinlicher macht. Biosynthese-Systeme des Sekundärmetabolismus sind allgemein ausgezeichnete Kandidaten für eine derartige Organisation, da sie für den Stoffwechsel des aufnehmenden Organismus zumindest anfänglich nicht unmittelbar essenziell sind, sondern statt dessen immer eine potenzielle Gewinnsituation darstellen, indem sie neue physiologische Funktionen verleihen (Dobrindt *et al.* 2004).

Daneben scheinen aber auch Übertragungen einzelner Gene vorzukommen, wie man in **Abb. 14** anhand der völlig vereinzelt stehenden Sequenz der AT-Domäne von NpF3156 aus *N. punctiforme* erkennen kann. Das Biosynthese-System, zu dem NpF3156 gehört, ist bislang biochemisch nicht charakterisiert worden, der produzierte Metabolit deshalb nicht bekannt. Bioninformatisch ließ sich der AT-Domäne allerdings mit hoher Wahrscheinlichkeit eine Spezifität für Methylmalonyl-CoA zuweisen. In diesem Fall kommt als Erklärung der phylogenetischen Verhältnisse ein horizontaler Transfer aus einem Myxobakterium in Frage. Andererseits fällt auf, dass es sich um die einzige Methylmalonyl-CoA-verwertende AT-Domäne aus Cyanobakterien im gesamten als Grundlage dienenden Datensatz handelt. Es ist daher nicht auszuschließen, dass sich die Situation mit der Zugänglichkeit weiterer cyanobakterieller PKS-I-Sequenzen verändert und ein eigenes AT-Cluster mit dieser Spezifität identifizierbar wird.

4.2.2.3 Evolutionärer Vergleich von PKS-I-Genclustern aus *Streptomyces*

Der im vorangegangenen Abschnitt erläuterte Transfer von Genen als Gesamtheit in Form von Genomischen Inseln findet sich auch in der Gruppe der actinobakteriellen *Streptomyces*. Dies wird

vermutlich durch die besondere Organisation des Genoms als lineares Chromosom erleichtert. Der Vergleich der beiden vollständig sequenzierten Genome von *Streptomyces avermitilis* und *Streptomyces coelicolor* zeigt, dass die Mehrzahl der Biosynthese-Cluster für Antibiotika oder allgemein für Metabolite des NRPS/PKS-Typs auf dem linearen Chromosom liegt und häufig von Transposase-Genen eingeschlossen wird (Bentley *et al.* 2002; Ikeda *et al.* 2003). Vereinzelte Hinweise auf konkrete HGT-Ereignisse, z. B. im Fall der Biosynthese-Gene für das Antibiotikum Streptomycin und diverse Sekundärmetabolite mit Polyketid-Anteilen, sind seit längerem bekannt (Egan *et al.* 1998; Wiener *et al.* 1998; Egan *et al.* 2001; Metsä-Ketelä *et al.* 2002). Besonders aufschlussreich war eine phylogenetische Analyse auf der Grundlage der Aminosäure-Sequenzen der KS- und AT-Domänen von 23 PKS-I-Clustern aus Streptomyceten (Ginolhac *et al.* 2005). Es konnte gezeigt werden, dass im Hinblick auf die ausgeprägte evolutionäre Verwandtschaft, die völlig identische Cluster-Konfiguration und die auffälligen GC-Gehalte das Vorkommen dreier Biosynthese-Cluster mit hoher Wahrscheinlichkeit auf HGT der gesamten Cluster-Gene zurückzuführen ist. Dabei handelt es sich um die Biosynthese-Cluster für Amphotericin, Nystatin und Pimaricin aus den Streptomyceten *S. nodosus*, *S. noursei* bzw. *S. natalensis*. Dies konnte in der vorliegenden Arbeit mit einem ähnlichen Analyseansatz reproduziert werden, wobei jedoch eine etwas veränderte Auswahl an Clustern getroffen wurde (siehe Abschnitt 3.3.2). In **Abb. 18** ist zu erkennen, dass die KS-Domänen der drei genannten Biosynthese-Systeme eine einzige phylogenetische Gruppe bilden, die sich von einem gemeinsamen Knoten ableitet. Auch die KS-Sequenzen aus dem Cluster der Candicidin-Biosynthese aus *Streptomyces* str. FR-008 fallen in diese Gruppe.

Darüber hinaus ergeben sich aus der Betrachtung der **Abb. 18** weitere interessante Aspekte. Im Gesamtbild verdienen die folgenden zwei Beobachtungen besondere Aufmerksamkeit. Zum einen ist deutlich erkennbar, dass die ganz überwiegende Mehrzahl von Sequenzen, die zum selben Gencluster gehören, in eine jeweils eigene phylogenetische Gruppe fallen, wie man es unter anderem im Fall der Gencluster von Oligomycin, Rifamycin und Spinosad sieht. Zum anderen gibt es jedoch auch Fälle von „vereinzelt“ KS-Domänen, die sich nicht von diesen Hauptästen ableiten lassen. In **Abb. 18** sind diese KS-Sequenzen aus dem Oligomycin-Cluster durch einzelne Sternchen gekennzeichnet. Die einfachste Erklärung für die beschriebene Monophylie von KS-Domänen desselben Biosynthese-Clusters wäre eine Reihe von sich ständig wiederholenden Genduplikationen, die von einem allen gemeinsamen Vorläufer ausgegangen wären. Das würde bedeuten, dass sich die Module eines Clusters – zumindest größtenteils – aus ein und demselben Anfangsmodul entwickelt hätten. Dieser Ablauf wiederum würde auch implizieren, dass diese Cluster praktisch vollständig in den einzelnen Streptomyceten-Spezies entstanden wären, also je eigene Entwicklungen darstellten. Diese Eigenständigkeit wäre auch insofern von Bedeutung, als selbst strukturell sehr ähnliche Polyketid-Verbindungen im Hinblick auf die Evolution der zugrunde liegenden Biosynthesen evolutionär unabhängig entstanden wären. Beispielsweise ergibt sich aus dem Baum in **Abb. 18** eine unverschränkte Evolution der Biosynthese-Cluster der beiden Makrolid-Polyketide Picromycin und Pimaricin. Somit würde gelten: Aus der strukturellen Ähnlichkeit zweier Polyketide lässt sich nicht unbedingt auf einen alle Komponenten der Biosynthese umfassenden gemeinsamen evolutionären Ursprung schließen.

Die Erklärung der zweiten Beobachtung, nämlich des Umstandes, dass manche KS-Domänen phylogenetisch außerhalb des Hauptclusters positioniert sind, macht nicht unbedingt die Annahme von

HGT erforderlich. Stattdessen kann diese Situation auf folgende Weise interpretiert werden: Die betreffenden Mitglieder der PKS-Multigen-Familie existierten bereits vor den Speziationsprozessen, die zur Diversität der rezenten Arten der Familie *Streptomycetaceae* führten. Das Verteilungsmuster der KS-Domänen – und der anderen Domärentypen – ergab sich dann aus einer so genannten „Birth and Death“-Evolution, wie sie mittlerweile für eine beachtliche Zahl von Multigen-Familien angenommen wird (Nei und Rooney 2005). Dieses Evolutionsmodell basiert auf der Annahme, dass bei der Entwicklung von Multigen-Familien nach den initialen Genduplikationen einige der Kopien für lange Zeit in der Entwicklungslinie erhalten bleiben, während andere inaktiviert und schließlich mittels Deletion aus dem Genom entfernt werden. In der Tat findet man z. B. im Genom von *S. avermitilis* PKS-artige Gene, die aufgrund von Mutationen wahrscheinlich nicht mehr funktionell sind und als fragmentierte Überreste einstmals funktionsfähiger Cluster interpretiert werden können (unveröffentlichte Daten).

Die phylogenetische Homogenität von Domänen eines Biosynthese-Clusters lässt sich, wie oben gesagt, als direkte Folge von wiederholten Genduplikationen interpretieren. Problematisch bei dieser Erklärung ist, dass man eher erwarten würde, im Zeitpunkt der Speziation hätten sich zumindest Teile der später voll ausgebildeten Cluster gewissermaßen *in statu nascendi* befunden. Dies würde allerdings zu anderen phylogenetischen Mustern führen. Hätte z. B. ein PKS-I-Cluster aus drei zusammenhängenden Modulen vor der Speziation existiert und wäre nach der vollzogenen Artbildung durch Duplikationen weiterentwickelt und vergrößert worden, so müsste die Phylogenie teilweise Überlappungen zwischen Sequenzen aus verschiedenen Clustern zeigen. Ein zur reinen Duplikationsabfolge alternativer Erklärungsansatz ist durchaus denkbar: Die fehlenden Überlappungen könnten durch das Phänomen der Genkonversion begründet sein. Unter Genkonversion versteht man eine nicht-reziprok ablaufende homologe Rekombination, in deren Verlauf eine Variante des entsprechenden DNA-Abschnitts verloren geht (Nei 2000). Sie ist die Basis für die so genannte konzertierte Evolution von Genfamilien, bei der sich die einzelnen der Genfamilie angehörenden Gene nicht unabhängig von den anderen Mitgliedern entwickeln, sondern einen hohen Grad an intrafamiliärer Sequenz-Homogenität erhalten. Eine Reihe theoretischer Untersuchungen konnte zeigen, dass Genkonversionseignisse zu einer konzertierten Evolution führen können (Birky und Skavaril 1976; Nagylaki und Petes 1982; Nagylaki 1984; Ohta 1990). Hinweise auf Genkonversionen wurden bereits in vielen Organismen für unterschiedliche Genfamilien gefunden. Bekannte Beispiele bei Eukaryoten umfassen die Genfamilien der α - und γ -Globine (Liebhaber, Goossens und Kan 1981; Scott *et al.* 1984), die „Heat Shock“-Proteine in *Drosophila*-Arten (Leigh Brown und Ish-Horowicz 1981) und die Gene des „Major Histocompatibility Complex“ (MHC), für die Genkonversion als Mechanismus zur Generierung von Polymorphismen vorgeschlagen wurde (Weiss *et al.* 1983a; Weiss *et al.* 1983b; Ohta 1998).

Im Bereich der Prokaryoten existieren ebenfalls viele gut untersuchte Fälle von Genkonversion, so unter anderem bei den Genen der ribosomalen RNA (Liao 2000), der Nitrogenase NifH in *Rhizobium etli* (Rodriguez und Romero 1998) und bei Genen, die für Proteine der äußeren Membran gramnegativer Bakterien codieren (Jordan *et al.* 2001). Für eine umfassende Übersicht sei an dieser Stelle auf einen kürzlich erschienenen Übersichtsartikel verwiesen (Santoyo und Romero 2005). Der entscheidende Effekt der Genkonversion besteht darin, dass die Sequenz-Homogenität der Mitglieder einer Genfamilie innerhalb einer Spezies weitaus größer ausfällt als zwischen zwei verschiedenen Spezies, auch wenn diese sehr eng verwandt sind. So besitzen die beiden γ -Proteobakterien *Escherichia coli* und *Haemophilus*

influenzae jeweils sieben Gene für die ribosomale 16S-RNA, für die aufgrund der sehr ähnlichen Anordnung ein gemeinsamer Ursprung unterstellt werden darf. Es handelt sich also mit großer Sicherheit um jeweils orthologe Paare. Dennoch findet man eine strikt Spezies-spezifische Clusterung im phylogenetischen Baum (Liao 2000). Ganz entsprechend könnte man die in **Abb. 18** dargestellte Situation interpretieren. Am Beispiel der beiden Biosynthese-Cluster für Oligomycin und Rifamycin hieße das: Es gab einen gemeinsamen Vorläufer-Cluster, aus dem sich aufgrund der Speziesbildung orthologe Sequenzpaare bildeten. Diese orthologen Paare, d. h. in Bezug auf den vorliegenden Datensatz orthologe Paare von KS-Domänen, müssten zwar jeweils auf gemeinsamen Zweigen des phylogenetischen Baums liegen. Dies wäre jedoch mittlerweile nicht mehr erkennbar, da durch diverse Genkonversionen die ursprünglichen KS-Varianten ersetzt und damit innerhalb eines Clusters homogenisiert worden wären.

Eine klare Entscheidung zwischen den beiden vorgestellten Erklärungsansätzen ist nicht möglich.

4.2.3 Biokombinatorik in der intragenomischen Evolution der PKS-I

Die in Abschnitt 3.3 vorgestellten Ergebnisse zur intragenomischen Evolution modularer PKS-I im Modellorganismus *Streptomyces avermitilis* lassen sich unter zwei verschiedenen Aspekten diskutieren. Von Interesse ist der Vergleich mit experimentellen Labormethoden, die den modularen Aufbau auszunutzen versuchen, um neue Verbindungen zu generieren. Zu diesem Aspekt gehört auch die Frage, wie man sich experimentell an die im Organismus ablaufenden Vorgänge annähern könnte. Der zweite Aspekt umfasst die allgemeine Bedeutung der modularen Struktur von Enzymsystemen für die Evolution von Proteinen.

4.2.3.1 Vergleich der natürlichen Biokombinatorik mit experimentellen Methoden und Ergebnissen

Schon seit längerem gab es Vermutungen über die evolutionären Abläufe, die zur Ausbildung der multimodularen Struktur von PKS-I führten. Ein Vorschlag lautete, dass ihre Evolution auf wiederholten Duplikationsrunden basiert, wodurch neue Module zum ursprünglichen System hinzukommen, und zwar entweder als Genfusionen oder in der Form neuer eigenständiger Proteine, die in den „Fließband-Prozess“ eingebunden werden (Hopwood 1997). Die vorhandene Vielfalt an unterschiedlich programmierten PKS-I sollte dabei auf den Austausch von Modulen zurückzuführen sein (Hopwood 1997). In der vorliegenden Arbeit wurde erstmals systematisch anhand des Genoms des Modellorganismus *S. avermitilis* untersucht, welche Arten von Substitutionen innerhalb von PKS-I-Systemen tatsächlich stattgefunden haben. Die Rekombinationsanalyse mittels vergleichender Gegenüberstellung der Phylogenien der einzelnen Domärentypen erlaubte den Nachweis folgender Prozesse:

- (1) Austausch von AT-Domänen unter Änderung der Substratspezifität (Abschnitt 3.3.4),
- (2) Einfügen bzw. Verlust einer KR-Domäne (Abschnitt 3.3.5) sowie
- (3) Austausch einer Einheit aus DH- und KR-Domäne gegen eine KR-Domäne (Abschnitt 3.3.5).

Nicht beobachtet wurde hingegen ein Austausch von KS-Domänen, was für biotechnologische Ansätze der Manipulation dieser Systeme von Interesse sein könnte. Dieser Umstand deutet darauf hin, dass die KS-Domänen eines Moduls mit anderen Komponenten auf eine besonders abgestimmte Weise kooperieren, so dass ihr Austausch mit einem hohen Ausfallrisiko einhergeht. Alle aufgelisteten Änderungen bedeuten eine Umwandlung eines bestimmten Modultyps in einen anderen und führen somit entsprechend dem Schema in **Abb. 5** zu einer Veränderung im synthetisierten Polyketid-Gerüst. Vergleicht man diese natürlich auftretenden biokombinatorischen Ereignisse mit publizierten Arbeiten zur experimentellen Biokombinatorik mit PKS des Typs I, so finden sich viele Parallelen. In einer Reihe von Arbeiten wurde der erfolgreiche Austausch einer AT-Domäne mit Spezifität für Methylmalonyl-CoA gegen eine solche mit Spezifität für Malonyl-CoA beschrieben (Oliynyk *et al.* 1996; Ruan *et al.* 1997; Lau *et al.* 1999). Außerdem konnten Ethyl-substituierte Erythromycin-Derivate erhalten werden, indem die natürlicherweise im Erythromycin-Biosynthese-Cluster vorhandene AT-Domäne, die Methylmalonyl-CoA umsetzt, gegen eine AT-Domäne ersetzt wurde, die das relativ seltene Ethylmalonyl-CoA als Substrat benutzt (Stassi *et al.* 1998).

Der oben genannte zweite Typ von Veränderung wurde im Effekt ebenfalls bereits im Labor erfolgreich durchgeführt. Das Entfernen der enzymatischen Ketoreduktase-Funktion gelang zunächst mittels verschiedener Arten der Inaktivierung von KR-Domänen. Diese wurden entweder durch zufällige PCR-basierte Mutagenese unbrauchbar gemacht oder durch natürlicherweise vorkommende nicht-funktionelle KR-Domänen ersetzt (Bedford *et al.* 1996). Später kamen auch echte Deletionen von KR-Bereichen hinzu. Um dies zu bewerkstelligen, wurde ein Sequenzabschnitt, der einen Teil des AT-KR-Linker-Bereichs, die KR-Domäne und ein Stück des KR-ACP-Linkers umfasste, durch eine nur 18 Aminosäure-Reste lange Sequenz ersetzt (McDaniel *et al.* 1999). Diese Vorgehensweise erinnert stark an die im Biosynthese-Cluster für Avermectin rekonstruierten Ereignisse (vgl. **Abb. 24**), abgesehen davon, dass in natürlichen PKS-I der Interdomänen-Abschnitt zwischen AT- und ACP-Domänen typischerweise ca. 90 Aminosäure-Reste umfasst (siehe **Abb. 5**).

Weiterhin wurde für die Erythromycin-Synthase der erfolgreiche Austausch eines kompletten Abschnitts aus DH- und KR-Domäne gegen eine einzelne KR-Domäne eines anderen Moduls berichtet (McDaniel *et al.* 1999), was zur Folge hatte, dass die Polyketid-Kette an der entsprechenden Position weiterhin eine Hydroxyl-Gruppe aufwies. Dies entspricht genau der dritten der oben aufgelisteten rekombinatorischen Modifikationen. Im Bezug auf Veränderungen unter Beteiligung von ER-Domänen war – wie im Ergebnisteil dargestellt – die Situation vergleichsweise unklar, da im Datensatz nur wenige Module inbegriffen waren, die diesen Domänentyp enthielten. Allerdings gab es Hinweise darauf, dass der Transfer großer Sequenzeinheiten unter Einschluss von DH-, ER- und KR-Domänen möglich zu sein scheint (siehe Abschnitt 3.3.5). Auch für diese Konstellation lassen sich Analogien aus der experimentellen Biokombinatorik mit PKS-I finden. Es war beispielsweise möglich, eine KR-Domäne durch eine komplette DH-ER-KR-Einheit zu ersetzen, was in einem Polyketid-Produkt mit voll abgesättigten, alkanischen C-Atomen anstatt der Hydroxyl-Gruppe des natürlichen Produktes resultierte (McDaniel *et al.* 1999).

Die aufgeführten Beispiele machen deutlich, dass es zwischen der Rekombinations-basierten Biokombinatorik, wie sie im Genom von *S. avermitilis* nachvollzogen werden kann, und den künstlichen Laboransätzen auf der phänomenologischen Ebene interessanterweise einige Berührungspunkte gibt. Die dahinter stehenden Mechanismen sind jedoch äußerst verschieden.

Experimentell bestand seit Beginn dieser Ansätze Mitte der 90er Jahre des 20. Jahrhunderts die Vorgehensweise grundsätzlich darin, durch Klonierungsarbeiten gezielt Sequenzabschnitte auszutauschen. Bedenkt man, wie aufwändig und fehlerträchtig jeder einzelne Schritt dabei üblicherweise ist, wird klar, dass sich dieses Verfahren nicht als echte Hochdurchsatz-Methode eignet. Eine der Schwierigkeiten besteht im Auffinden geeigneter Restriktionsstellen, die – falls nicht vorhanden – zu allererst durch einzelne Basenaustausche generiert werden müssen (Reeves 2003). Außerdem wurden immer wieder Inkompatibilitäten zwischen neu zusammengefügt Domänen, die aus ihrem ursprünglichen Zusammenhang entfernt wurden, beobachtet (Reeves 2003).

Wollte man die Natur bei der Generierung von Polyketid-Substanzbibliotheken wirklich nachahmen, müsste man versuchen, homologe Rekombination als Basis des Verfahrens zu etablieren. Eine Methode, die in diese Richtung zielt, stellt das so genannte „DNA shuffling“ dar. Die Idee besteht darin, Varianten verwandter Gene zufällig zu fragmentieren und dann diese Fragmente mittels Überhängen, die durch eine PCR angefügt werden, neu miteinander zu kombinieren (Patten, Howard und Stemmer 1997). Theoretisch wäre es auf diese Weise möglich, eine sehr große Zahl neuer Strukturvarianten bekannter Polyketide zu produzieren. Allerdings bliebe das Hauptproblem aller *in vitro*-Verfahren bestehen. Da in solchen Ansätzen keine biologische Selektion wirksam wird, ist davon auszugehen, dass die zahlreichen nicht-funktionellen Neukombinationen die sehr wenigen erfolgreichen überdecken würden. Zu bedenken ist hier auch die bei Neukombination von Modulen eines PKS-Systems häufig zu beobachtende geringe Ausbeute an Produkt.

Kürzlich wurde eine neue Methode zur Generierung von Substanzbibliotheken auf Basis von modularen PKS des Typs I vorgestellt (Menzella *et al.* 2005), die sich den natürlichen biokombinatorischen Abläufen deutlich mehr annähert als alle bisherigen Ansätze. Das Verfahren benutzt zwei Typen von Modulen, Beladungsmodule und Elongationsmodule. Diese Modultypen werden jeweils auf eigenen Plasmiden codiert bereitgestellt und in einen *E. coli*-Stamm zur heterologen Expression eingebracht. Besondere Beachtung fand die Erkenntnis, dass die *linker*-Regionen zwischen den Domänen von großer Bedeutung für eine funktionelle Zusammenarbeit zwischen den PKS-Einheiten sind (siehe z. B. (Gokhale *et al.* 1999)). Die wesentliche Neuerung des Verfahrens besteht darin, eine universelle *linker*-Technologie mit einer schnellen und effizienten Synthese langer künstlicher Gene (Kodumal *et al.* 2004) zu verbinden. Diese Technik ermöglicht es, die gewünschten Genkassetten unabhängig von der Bakteriengruppe, aus der sie stammen, in das System zu integrieren, da eine Anpassung an den Codon-Gebrauch von *E. coli* erfolgt. Außerdem können gleichzeitig auf einfache Weise passende Schnittstellen für Restriktionsenzyme eingeführt werden. Diese Restriktionsstellen flankieren Domänen, *linker*-Bereiche und ganze Module, so dass leicht austauschbare Bausteine entstehen. Obwohl also der Austausch von Domänen oder Domänenabschnitten gemäß der Konzeption des Verfahrens möglich wäre, beschränkte sich die experimentelle Evaluierung bislang darauf, neue Kombinationen vollständiger Module herzustellen (Menzella *et al.* 2005). Es wäre interessant, diese Methode auch auf einzelne Domänen und Abschnitte innerhalb eines Moduls anzuwenden, also den Aufbau der Module durch Neukombinationen zu verändern, da diese Vorgehensweise den im Genom von *S. avermitilis* detektierbaren Prozessen entspräche.

Auch diese methodische Neuentwicklung ändert jedoch nichts an dem grundsätzlichen Unterschied im Ablauf natürlicher und experimenteller Biokombinatorik. Während das Bakterium auf seine Rekombinationsmaschinerie zurückgreift, stützen sich die diversen Labormethoden auf Restriktionsverdau und Religation. Im Prinzip sollte es möglich sein, einen experimentellen Ansatz auszuarbeiten, der

tatsächlich auf Rekombinationsprozessen basiert. Studien zur experimentellen Rekombination laufen häufig nach folgendem Schema ab: Die ausgewählten Allele werden im Organismus durch Mutationen unbrauchbar gemacht und dann mit Hilfe von unbeschädigten Kopien homologer Gene repariert (siehe zu diesem Thema die beiden Übersichtsartikel (Santoyo und Romero 2005; Volodin, Voloshin und Camerini-Otero 2005) und die darin genannten Literaturstellen). Allerdings konnten Genkonversionen in pathogenen Bakterienstämmen, die zur Diversifizierung von Oberflächen-Antigenen beitragen, auch auf andere Weise nachgewiesen werden (Santoyo und Romero 2005; Volodin, Voloshin und Camerini-Otero 2005). Die Schwierigkeit hinsichtlich von Rekombinationsereignissen in modularen PKS-I besteht vermutlich darin, dass die Häufigkeit dieser Ereignisse deutlich geringer ausfällt als z. B. bei der Variation von bakteriellen Oberflächen-Antigenen. Um den Einfluss der Rekombination in sinnvollen Zeiträumen erfassen zu können, wäre deshalb das Wirken eines Selektionsdrucks vonnöten. Das Produkt des in Frage stehenden PKS-Systems müsste also unter den Versuchsbedingungen dem produzierenden Bakterium einen Vorteil verschaffen, der eine leichte Selektion erlaubt. Ausgangspunkt des Experiments wäre das Einbringen einer zur Nicht-Funktionalität führenden Mutation in das PKS-System. Wenn der Stamm weitere PKS-I enthält, sollte er in der Lage sein, die beschädigten Abschnitte, z. B. Module, innerhalb des Systems zu reparieren. Zur Erhöhung der Rekombinationsfrequenz könnten solche mutierten Stämme eingesetzt werden, die nicht mehr über ein funktionsfähiges *mismatch*-Reparatursystem verfügen (Hranueli *et al.* 2005). Mit einem solchen experimentellen Ansatz könnten nicht nur die Möglichkeiten und Grenzen bei der Rekombination von PKS-Genen erforscht werden, sondern auch Fragen nach der Reziprozität oder Nicht-Reziprozität der Austauschprozesse.

4.2.3.2 Evolutionäre Bedeutung der modularen Enzymorganisation: Modularität und Evolvierbarkeit

Multienzymkomplexe, wie sie die PKS vom Typ I darstellen, eröffnen einem Organismus besondere Möglichkeiten zur Fortentwicklung seines Stoffwechsels. Innerhalb der Themenkomplexe Proteinevolution und Generierung metabolischer Diversität verweisen die in den PKS-I von *S. avermitilis* aufgezeigten Rekombinationsereignisse auf die Prinzipien der Modularität und Evolvierbarkeit. Im Folgenden wird diskutiert, inwieweit die spezielle Architektur und die sich daraus begründende „Plastizität“ der modularen PKS-I gerade im Zusammenhang des Sekundärmetabolismus von Bedeutung sind.

Es besteht kein Zweifel mehr daran, dass die Duplikation von Genen einen wesentlichen Bestandteil der Evolution neuer Proteinfunktionen darstellt. Diese Hypothese wurde bereits vor knapp vier Jahrzehnten formuliert (Ohno 1970) und in jüngerer Zeit um neue Konzepte bereichert. Das ursprüngliche Modell geht davon aus, dass der erste Schritt in der Ausbildung einer Genkopie besteht, die gewissermaßen von ihren Aufgaben „freigestellt“ wird und zumindest anfänglich praktisch keinem Selektionsdruck unterliegt. Dieser Ansatz ist jedoch insbesondere in einem Punkt recht problematisch. Die Mutationen in dieser Genkopie sind nämlich in den allermeisten Fällen funktionell und strukturell nachteilig für das entsprechende Protein, weshalb die Wahrscheinlichkeit, auf diese Weise eine neue katalytische Funktion zu generieren, selbst nach evolutionsbiologischen Maßstäben außerordentlich gering wäre (Hughes 1994). Seit einigen Jahren wird jedoch wieder verstärkt auf das Phänomen der so genannten „Enzym-Promiskuität“ (englisch: *enzyme promiscuity*) hingewiesen (siehe hierzu z. B. die beiden

Übersichtsartikel (Copley 2003; Khersonsky, Roodveldt und Tawfik 2006)). Unter „Enzym-Promiskuität“ versteht man die Fähigkeit eines Enzyms, am aktiven Zentrum neben der Hauptreaktion (Primärreaktion), die dem Enzym zu eigen ist, auch Nebenreaktionen (Sekundärreaktionen) durchzuführen, wenngleich letztere typischerweise mit deutlich geringerer katalytischer Effizienz (O'Brien und Herschlag 1999). Ebenfalls recht frühzeitig wurde der Gedanke geäußert, dass der tatsächliche Ausgangspunkt der Evolution neuer Enzymfunktionen in solchen Mehrfachreaktionen liegen könnte (Jensen 1976). Wenn sich solche Alternativreaktionen herausgebildet hätten, würden sie nach der Duplikation eine neue enzymatische Aktivität darstellen, die möglicherweise bereits nahe am „Threshold“ eines selektiven Vorteils läge und somit weit höhere Chancen hätte, tatsächlich im Genom fixiert zu werden (O'Brien und Herschlag 1999). Im vorliegenden Zusammenhang kommt es auf folgendes an: Beide Modelle verdeutlichen, dass die Neuentwicklung und Verankerung innovativer enzymatischer Aktivitäten mittels Veränderungen im aktiven Zentrum einen langwierigen Prozess darstellen mit einer insgesamt geringen Wahrscheinlichkeit für ein positives Ergebnis, verbunden mit der prinzipiellen Schwierigkeit, dass dieser Weg der Proteinevolution immer nur einen extrem kleinen Bruchteil des theoretisch zur Verfügung stehenden Sequenzraums durchprobiert (Griffiths und Tawfik 2000).

Die modular gebauten PKS des Typs I zeigen – ebenso wie die auch in Modulen organisierten NRPS – einen anderen Weg der Enzymevolution auf, bei dem nach dem Baukastenprinzip die Grundbausteine neu kombiniert werden können. Bildlich sehr treffend wurde dieses Prinzip als *Lego-ization* bezeichnet (Sherman 2005). Dadurch wird es möglich, in kurzer Zeit eine große Zahl verschiedener Verbindungen zu synthetisieren. Die Wirksamkeit dieses Prinzips wurde bereits mehrfach nachgewiesen. Mit einer phylogenetischen Analyse des Biosynthese-Clusters für Microcystin in verschiedenen Stämmen des Cyanobakteriums *Microcystis* konnte beispielsweise gezeigt werden, dass Rekombinationsereignisse an ihrer Evolution beteiligt waren (Tanabe, Kaya und Watanabe 2004). Sequenzanalysen ergaben auch eine hohe Wahrscheinlichkeit für Rekombinationsereignisse zwischen sich funktionell entsprechenden Domänen zweier Module der Microcystin-Synthetase, die sich mit den jeweiligen Profilen an Microcystin-Isoformen in diesen *Microcystis*-Stämmen korrelieren lassen (Mikalsen *et al.* 2003). Die große Bedeutung der homologen Rekombination wurde auch für andere Proteintypen beschrieben. So basiert die hohe Variabilität von Oberflächen-Antigenen pathogener Bakterien, welche die immunologische Abwehr und Behandlung dieser Keime erschweren, auf homologer Rekombination. Unter anderem wurde dies für die Variation von Pilinen in *Neisseriae* (Kline *et al.* 2003), die Lipoprotein-Diversität in *Borrelia burgdorferi* (Zhang *et al.* 1997) und die auffällige Heterogenität in der Sequenz des so genannten „Hauptantigens“ von *Treponema pallidum* (Centurion-Lara *et al.* 2004) gezeigt. In diesen Fällen fußt die Variabilität auf einer Reihe von Genkassetten, die „still“ im Genom liegen und mittels Rekombination in den Genabschnitt eingebracht werden können, der von dem Bakterium tatsächlich expremiert wird. Die Generierung neuer Proteinkompositionen mittels Rekombination findet sich auch bei *Response*-Regulatoren bakterieller Systeme zur Signaltransduktion (Pao und Saier 1995).

Die Modularität von PKS-I und NRPS zeigt sich funktionell in zweierlei Hinsicht (Khosla und Harbury 2001). Die Erkennung und Bindung des elektrophilen und nucleophilen Reaktionspartners ist auf zwei unterschiedlichen Domänen lokalisiert, nämlich bei den PKS auf der KS- bzw. AT-Domäne, bei den NRPS auf der C- bzw. A-Domäne. Hierdurch werden diese beiden Funktionen separierbar und neu kombinierbar. Das zweite modulare Charakteristikum ist das *Channelling* innerhalb des Multienzym-

Komplexes: Die Biosynthese-Intermediate werden nicht mittels Diffusion an die nachfolgende Aktivität weitergegeben, sondern sind kovalent gebunden und werden direkt von einem aktiven Zentrum zum nächsten transferiert, ohne dass es zu Gleichgewichtseinstellungen mit der Flüssigkeitsumgebung kommt (Geck und Kirsch 1999). Die Limitierung des Prinzips der Modularität scheint darin zu bestehen, dass die ständigen Neukombinationen keine weitreichende katalytische Optimierung erlauben, wie sie bei vielen Enzymen mit ausgeprägter Substratspezifität zu beobachten sind. Aus Untersuchungen der Katalyse der Bindungsbildung in PKS-I und NRPS ist der Einfluss der Substrate auf die katalytische Effizienz der Umsetzung bekannt (Belshaw, Walsh und Stachelhaus 1999; Wu *et al.* 2000b). Im Allgemeinen sind die Umsetzungsraten auch nicht besonders hoch. Gerade im Zusammenhang des Sekundärmetabolismus scheint dies aber von untergeordneter Bedeutung zu sein. Die Möglichkeit, enzymatische Aktivitäten auf einfache Weise neu mischen und abstimmen zu können, ist offensichtlich ein großer evolutionärer Vorteil. Dabei kommt es in den Domänen nicht zu Sequenzanpassungen mittels Mutationen, die Aminosäurereste verändern, und nachfolgender positiver Selektion. Die Selektionsanalyse für die PKS-I aus *S. avermitilis* ergab vielmehr, dass ausschließlich negative Selektion (*purifying selection*) zur Absicherung von Struktur und Funktion zu beobachten ist.

Die modulare Architektur und Funktionsweise fügt sich darüber hinaus auch sehr gut in das Konzept der Evolvierbarkeit (englisch: *evolvability*) ein. Unter Evolvierbarkeit versteht man im weitesten Sinne die Fähigkeit eines biologischen Systems zur Evolution. Im engeren Sinne bezeichnet dieser Terminus jedoch eine besonders ausgeprägte Fähigkeit eines Organismus, vererbare phänotypische Variation hervorzubringen (Kirschner und Gerhart 1998). Ursprünglich wurde dieses Konzept in der Evolutionsbiologie entwickelt, um die „Entwickelbarkeit“ bestimmter morphologischer Eigenschaften besser beschreiben zu können. Es kann jedoch auch auf molekulare und zelluläre Prozesse übertragen werden. Insbesondere können auch quantitative und theoretische Modelle entwickelt werden (siehe z. B. (Conrad 1990)). Betont werden muss, dass Evolvierbarkeit im hier interessierenden Sinn nicht einfach die Beobachtung von Veränderungen meint, sondern quantitativ interpretiert die Wahrscheinlichkeit, mit der die Veränderungen auftreten, zum Ausdruck bringt. Die Fähigkeit zur Evolution lässt sich in zwei Komponenten untergliedern, die der evolutionären Weiterentwicklung förderlich sind: (1) Reduzierung der potenziellen Schädlichkeit (im Extrem: Letalität) von Mutationen und (2) Reduzierung der Zahl an Mutationen, die zur Ausbildung neuer nützlicher Eigenschaften notwendig sind (Kirschner und Gerhart 1998). In der Biologie sind eine Reihe von Systemen bekannt, die eine besonders gut ausgeprägte Evolvierbarkeit aufweisen. Ein gut untersuchtes Beispiel sind die Immunglobuline, deren Gene so organisiert sind, dass die Mutationsraten in den Komplementarität-bestimmenden Bereichen deutlich höher sind als in den Regionen, die für die einheitlich konstante Grundstruktur der Immunglobuline codieren. Dadurch wird die Wahrscheinlichkeit, nützliche Neuvarianten zu erzeugen, vergrößert (Kepler 1997; Plotkin und Dushoff 2003).

In biologischen Systemen besteht ein prinzipieller Zusammenhang zwischen modularer Organisation und Evolvierbarkeit. In den letzten Jahren ist dieses Thema gerade in der theoretischen Biologie intensiv diskutiert worden (siehe z. B. (Rainey und Cooper 2004) und die dort genannten Literaturstellen). Modulare PKS-I zeigen genau die Eigenschaften, die von einem gut evolvierbaren System zu erwarten sind: Durch ihre Struktur wird die Austauschbarkeit über homologe Rekombination erleichtert und somit die Gefahr, funktionslose Systeme zu schaffen, verringert; und die Zahl der

notwendigen Mutationen zur Ausbildung einer neuen Funktion, d. h. einer neuen synthetisierten Struktur, ist klein, da im Prinzip bereits ein Rekombinationsereignis ausreicht, um Veränderungen am produzierten Polyketid hervorzurufen.

4.3 PKS- und NRPS-Biosynthese-Cluster in *Nostoc punctiforme*

Die Kenntnis der genetischen Struktur aller PKS- und NRPS-Biosynthese-Cluster innerhalb eines Genoms ist aus einer Reihe von Gründen wichtig. Sie bildet beispielsweise die Basis von Rekombinationsanalysen, wie sie für den Modellorganismus *S. avermitilis* in dieser Arbeit durchgeführt wurden (siehe Abschnitte 3.3 und 4.2.3). Auch vergleichende Untersuchungen zwischen den unterschiedlichen Bakteriengruppen können auf dieser Grundlage besonders aussagefähig werden, etwa im Fall von Genclustern aus Cyanobakterien und Myxobakterien, für die sich teilweise enge phylogenetische Beziehungen ableiten ließen (siehe Abschnitt 3.2.3). Mittels der veröffentlichten vollständigen Genom-Sequenz von *Nostoc punctiforme* konnten mittels eines so genannten *genome mining*-Ansatzes Biosynthese-Cluster des PKS- und NRPS-Typs in diesem Cyanobakterium identifiziert werden. Dieser Ansatz wird im Folgenden in einen allgemeineren Zusammenhang eingeordnet, und es werden anhand der erhaltenen Ergebnisse auch mögliche Schwierigkeiten bei der Durchführung diskutiert.

4.3.1 *Genome mining* im Cyanobakterium *Nostoc punctiforme*

Durch die rasante Zunahme der Zahl vollständig sequenzierter Genome, insbesondere von prokaryotischen Organismen mit ihren vergleichsweise kleinen Genomen, hat sich die Verfahrensweise zum Auffinden neuer Sekundärmetabolite wesentlich gewandelt. Die Forschung in diesem Bereich wurde in der Folge bereits in vielerlei Hinsicht methodisch verändert, wobei diese Entwicklung auch in der Zukunft anhalten wird (Bode und Müller 2005). Die klassische Reihenfolge der Analyseschritte bei der Suche nach Metaboliten kann folgendermaßen beschrieben werden: Zunächst werden die Verbindungen – etwa im Verlauf von Screening-Verfahren – identifiziert und anschließend aufgereinigt und weiter charakterisiert. Im Anschluss daran wird versucht, die Gene des Biosynthese-Clusters in dem betreffenden Organismus zu identifizieren. Das Auffinden potenzieller Gencluster, die für PKS oder NRPS (oder für andere Biosynthese-Systeme des Sekundärstoffwechsels) codieren, wurde mit den methodischen Fortschritten der Molekulargenetik immer einfacher. Vor allem die Möglichkeit, relativ schnell und kostengünstig Genbibliotheken von Bakteriengenomen anzulegen, stellt ein wichtiges Mittel zur Charakterisierung solcher Cluster dar. Steht hingegen, wie heute oftmals der Fall, die gesamte Genomsequenz eines Organismus zur Verfügung, so lässt sich diese unter einer spezifizierten Fragestellung gezielt mit Hilfe bioinformatischer Verfahren auswerten. Für diese Herangehensweise wurde in der englischsprachigen Literatur in Anlehnung an den Begriff *data mining*, der allgemein die gezielte Datensuche und –auswertung beschreibt, die Bezeichnung *genome mining* geprägt. Alternativ spricht man auch von *genome screening*-Verfahren.

Von *genome mining* wird also immer dann gesprochen, wenn man sich bekannte Sequenzdaten von Genomen zunutze macht, um bestimmte „Zielobjekte“ in einem Organismus aufzufinden, wobei der Bogen der Anwendungen und die Kombinierbarkeit mit anderen Methoden weit gespannt sind. Dieses Verfahren wird mittlerweile häufig eingesetzt, um noch unbekannte Gene und Gencluster des Sekundärstoffwechsels zu identifizieren (Wilkinson und Micklefield 2007). Auf diese Weise konnten z. B. neue von PKS/NRPS-Systemen synthetisierte Verbindungen in dem Pilz *Aspergillus nidulans* entdeckt (Bergmann *et al.* 2007) und ein neuer Peptid-Metabolit aus dem bereits sehr gut untersuchten Actinobakterium *Streptomyces coelicolor* charakterisiert werden (Lautru *et al.* 2005).

In Anlehnung an das Verfahren des *genome mining* wurde die genomische Sequenz von *N. punctiforme* auf die Existenz von PKS- und NRPS-Genclustern durchsucht. Wie in Abschnitt 3.4 dargestellt, konnten elf Gencluster gefunden werden (siehe **Tab. 11**), von denen wiederum nur ein einziges Cluster mit einem bereits beschriebenen Metaboliten in Verbindung gebracht werden konnte, nämlich das Nostopeptolid-Cluster (Cluster Nr. IV in der hier benutzten Zählung). Aus den Cluster-Darstellungen der **Abb. 28** bis **Abb. 38** geht hervor, dass sieben dieser Biosynthese-Cluster einen ABC-Transporter aufweisen und dass in unmittelbarer Nähe von sechs Clustern Transposase- oder Integrase-Gene vorliegen. Die Assoziation mit ABC-Transportern legt die Möglichkeit eines Exportes der synthetisierten Verbindung nach außen nahe, wobei die Transportraten von der Stoffwechselsituation oder bestimmten anderen Stimuli abhängen können. Transposase- und Integrase-Gene verweisen darauf, dass der entsprechende Genabschnitt Teil eines mobilen genetischen Elements gewesen sein und in den Organismus eingeschleust worden sein könnte (siehe hierzu auch Abschnitt 4.2.2.2).

Neben dem einfachen Auffinden bestimmter Enzymfamilien mittels BLAST- oder Homologiesuchen macht sich das *genome mining* auch weiter gehende Analysemöglichkeiten zunutze. Im Fall von PKS-I und NRPS etwa können mit einer bestimmten Wahrscheinlichkeit die Substratspezifitäten der AT- bzw. der A-Domänen bestimmt werden. Die Anwendung von Struktur-basierten Modellen der Substraterkennung von A-Domänen (siehe Stachelhaus, Mootz und Marahiel 1999; Challis, Ravel und Townsend 2000) und Abschnitt 1.3.3 der Einleitung) ermöglichte die Vorhersage der vermutlich in die Metaboliten eingebauten Aminosäure-Reste (siehe **Abb. 28** bis **Abb. 38** im Ergebnisteil). Diese Prognosen konnten dann mit den experimentellen Ergebnissen (**Tab. 12**) verglichen werden, was in zwei Fällen eine eindeutige Zuordnung brachte. Zum einen ließ sich auf diesem Wege das bekannte Nostopeptolid (in der Variante A1 oder A3) und eine Gruppe von vermeintlichen Nostopeptolid-Derivaten Signalen im Chromatogramm zuordnen. Zum anderen konnte der Peak Nr. 4 in **Abb. 39** aufgrund der exakt passenden Aminosäure-Reste und einer stimmigen Gesamtmasse mit größter Sicherheit dem Biosynthese-Cluster V (siehe **Tab. 11** und **Abb. 32**) zugewiesen werden. Für dieses reine NRPS-Cluster konnte somit der *genome mining*-Ansatz konsequent bis zur Identifizierung der Substanz durchgeführt werden. Auf diesen Vorarbeiten aufbauend, wird es möglich sein, die Chromatographie-Bedingungen so zu verändern, dass die Auftrennung im entsprechenden Bereich optimiert wird, um den Peptid-Metaboliten zu isolieren und bei Vorliegen ausreichender Mengen strukturell zu charakterisieren.

Obwohl das Konzept des *genome mining* im Fall des Clusters V also erfolgreich angewendet werden konnte, zeigen die vorgelegten Ergebnisse zugleich seine Beschränkungen. Von den elf im Genom identifizierten Biosynthese-Clustern konnten im Rahmen des Metabolit-Screenings nur zwei Cluster sicher und ein Cluster mit nicht auszuräumenden Unsicherheiten zugewiesen werden. Dies

bedeutet, dass für die acht verbleibenden Cluster zunächst keine Zuordnung möglich war, sie also im Metabolit-Profil nicht gefunden werden konnten. Selbst wenn man Optimierungsmöglichkeiten in der Methodik der Profilierung mittels HPLC und Massenspektrometrie annimmt, ist davon auszugehen, dass die Zuweisungen einzelner Gencluster zu bestimmten Verbindungen lückenhaft bleiben werden. Der Grund hierfür ist das Phänomen der „stillen Gencluster“, die unter vielen Kulturbedingungen gar nicht oder nur sehr schwach expremiert werden, so dass die zugeordneten Metabolite nicht nachweisbar sind (Bode und Müller 2005). Bedenkt man die möglichen Funktionen von Sekundärmetaboliten als Signal- oder Abwehrstoffe, ist zu erwarten, dass viele von ihnen sehr selektiv nur unter ganz bestimmten Umgebungsbedingungen oder auf sehr spezifische Stimuli hin produziert werden. Daher rührt auch das so genannte „one strain – many compounds“-Konzept, abgekürzt mit dem Akronym OSMAC, wonach ausgiebige Variation der Anzuchtbedingungen, z. B. auch der Zusatz von organischen Lösungsmitteln, bei ein und demselben Bakterienstamm zum Auftreten neuer Sekundärmetabolite führt (Bode *et al.* 2002). Beim *Screening* der *N. punctiforme*-Kulturen wurden dementsprechend unterschiedliche Kulturbedingungen eingesetzt (u. a. Nährstoffbedingungen wie Stickstoff- oder Phosphor-Limitierung, Lichtverhältnisse und Temperatur), allerdings konnten damit keine neuen Signale in den HPLC-Messungen erhalten werden. Dies kann ein Hinweis darauf sein, dass manche dieser Metabolite selbst nach – eventuell durch Probieren gefundener – Induktion nur in sehr geringen Mengen hergestellt und deshalb in einer typischen Profilierung mittels HPLC übersehen werden. In solchen Fällen kann im Prinzip Abhilfe geschaffen werden, indem man die Transkription des dahinter stehenden Synthese-Apparates künstlich aktiviert (Metsä-Ketelä, Ylihonko und Mantsala 2004).

4.3.2 Isolierung eines möglichen Metaboliten des Biosynthese-Clusters VII

In den HPLC-Chromatogramm wurde auch eine Reihe von Absorptionssignalen sichtbar, die bei der massenspektrometrischen Untersuchung mittels MALDI-TOF keine hinreichenden Signalstärken erreichten und daher nicht sicher bestimmbar waren. Um die aus dem *genome mining* erhaltenen Cluster dennoch zuordnen zu können, kann die klassische Strategie angewandt werden, von ausgewählten Clustern Knockout-Mutanten herzustellen, die den Metaboliten nicht mehr zu produzieren vermögen. Im Vergleich zum Wildtyp-Stamm sollten dann im HPLC-Profil Unterschiede sichtbar werden, falls der Metabolit tatsächlich unter den betrachteten Bedingungen hergestellt wird und nachweisbar ist. Wie in Abschnitt 3.6.3 dargestellt, gelang dies bei *N. punctiforme* für das Cluster VII, einem Biosynthese-Cluster, das vorwiegend aus PKS-Modulen aufgebaut ist, jedoch daneben auch eine einmodulare NRPS-Komponente besitzt. Eben dieses NRPS-Modul (NpF3173), dessen Funktionstüchtigkeit nach heterologer Expression in einem Enzym-Assay nachgewiesen werden konnte (siehe Abschnitt 3.6.2.2), wurde durch Integration einer Kanamycin-Resistenzkassette in den Leserahmen unbrauchbar gemacht. Während in den Extrakten von Zellpellets keine deutlichen Unterschiede festzustellen waren, hatte der Knockout von NpF3173 im Mediumüberstand zum Ergebnis, dass ein bei Wildtyp-Kulturen registrierbarer Peak völlig fehlte (siehe **Abb. 50**). Das Auftreten dieser Verbindung im Kulturüberstand bei vergleichsweise geringen Mengen im Extrakt des Zellpellets (siehe **Abb. 39**) steht in guter Übereinstimmung mit der Konfiguration dieses Biosynthese-Clusters. Das Cluster VII codiert nämlich, wie in **Abb. 34** dargestellt, auch für die Komponenten eines ABC-Transporters, weshalb ein Export der synthetisierten Verbindung durchaus den Erwartungen entspricht.

Aus der Cluster-Struktur lassen sich auch Erwartungen über bestimmte Eigenschaften des Metaboliten ableiten. Geht man davon aus, dass alle Module funktionell und aktiv sind, ist ein Molekulargewicht von etwa 450 zu erwarten. Diese Schätzung ergibt sich aus dem Einbau von neun Acetat-Einheiten der entsprechenden Reduktionsstufen und eines Cystein-Restes, der zum Thiazolin cyclisiert wird. Weiterhin ist aufgrund des Thiazolin-Ringes Schwefel als Bestandteil zu erwarten. Eine massenspektrometrische Analyse dieser Verbindung ergab einen Masse-Ladungs-Quotienten von $m/z=252$. Die extrem genaue Massenbestimmung mittels FT-ICR-MS führte zur Summenformel $C_{15}H_{24}O_3$, wobei mit absoluter Sicherheit ausgeschlossen werden konnte, dass Schwefel in der Verbindung enthalten ist. Diese Ergebnisse decken sich also nicht mit den aus der Modul-Konfiguration zu erwartenden Eigenschaften. Hierfür gibt es mehrere Erklärungen. Zunächst wäre es möglich, dass die identifizierte Verbindung nicht von Cluster VII synthetisiert wird, ihre Herstellung jedoch bei einem Ausfall dieses Clusters ebenfalls tangiert wird. Für einen zweiten Erklärungsansatz muss berücksichtigt werden, dass eine exakte Struktur- oder auch nur Massenvorhersage ausschließlich auf Basis der Kenntnis des Genclusters häufig nicht möglich ist, da sich unter anderem nicht sicher ableiten lässt, (1) ob Module wirklich funktionell bzw. aktiv sind oder ausgelassen werden (so genanntes *skipping*), (2) welche Bausteine tatsächlich als Substrate dienen und (3) inwiefern andere Enzyme an der Biosynthese Anteil haben, die nicht Bestandteil des Clusters sind, sondern in andere Teilen des Genoms codiert werden.

Im Fall des hier betrachteten Biosynthese-Clusters VII fällt im Vergleich zu den anderen Clustern eine gewisse Fragmentiertheit ins Auge. So bilden z. B. die verschiedenen PKS-Module keine durchgehende Abfolge, sondern sind durch einige andere offene Leserahmen voneinander getrennt, weshalb in der Tat die Frage gestellt werden muss, ob alle Module an der Biosynthese beteiligt sind. Aus der gegebenen Cluster-Struktur kann eine Biosynthese eines Metaboliten mit einem Molekulargewicht von 252 konstruiert werden, wenn man folgende Annahmen zulässt: (1) das NRPS-Modul ist nicht aktiv, das erste Modul von NpF3169, dessen KS-Domänen Ähnlichkeiten mit KS-Domänen in Beladungsmodulen hat, dient als Beladungsmodul und (3) das Modul NpF3158 ist ebenfalls nicht aktiv. Daraus ergäbe sich das in **Abb. 55** dargestellte Schema. An dieser Stelle sei betont, dass dieser Vorschlag einen sehr spekulativen Charakter hat. Er dient hauptsächlich dem Zweck aufzuzeigen, dass die experimentell gefundenen Eigenschaften in nicht völligem Widerspruch zur genetischen Struktur des Clusters stehen.

Die Hauptfrage in diesem Zusammenhang lautet: Wie ist zu erklären, dass der aus der Aminosäure Cystein zu erwartende Molekülanteil durch exakte Massenbestimmung nicht nachweisbar ist und dennoch die Biosynthese des Metaboliten durch einen Knockout, der die NRPS-Komponente betrifft, unterbunden werden kann? Zur weiteren Abklärung dieses Sachverhalts muss der Metabolit in größeren Mengen aufgereinigt werden, um mittels anschließender NMR-Analyse Anhaltspunkte zur Struktur zu erhalten.

Weiterhin kann die hergestellte Mutante – ergänzt um weitere Knockout-Konstrukte – dazu herangezogen werden, die möglichen physiologischen Funktionen des Metaboliten zu untersuchen. Eine Schwierigkeit bei der Beurteilung der Evolution von PKS und NRPS und allgemein des Sekundärstoffwechsels besteht gerade darin, dass für viele der beschriebenen Metabolite noch keine Funktion beschrieben ist, die im physiologischen und ökologischen Zusammenhang des produzierenden Organismus von Bedeutung sein könnte. Solche Kenntnisse sind jedoch notwendig, um die evolutionären Mechanismen vertieft herauszuarbeiten, die der Generierung der extremen Strukturvariabilität zugrunde liegen, und damit allgemein gültige Theorien zur Evolution des Sekundärstoffwechsels zu formulieren.

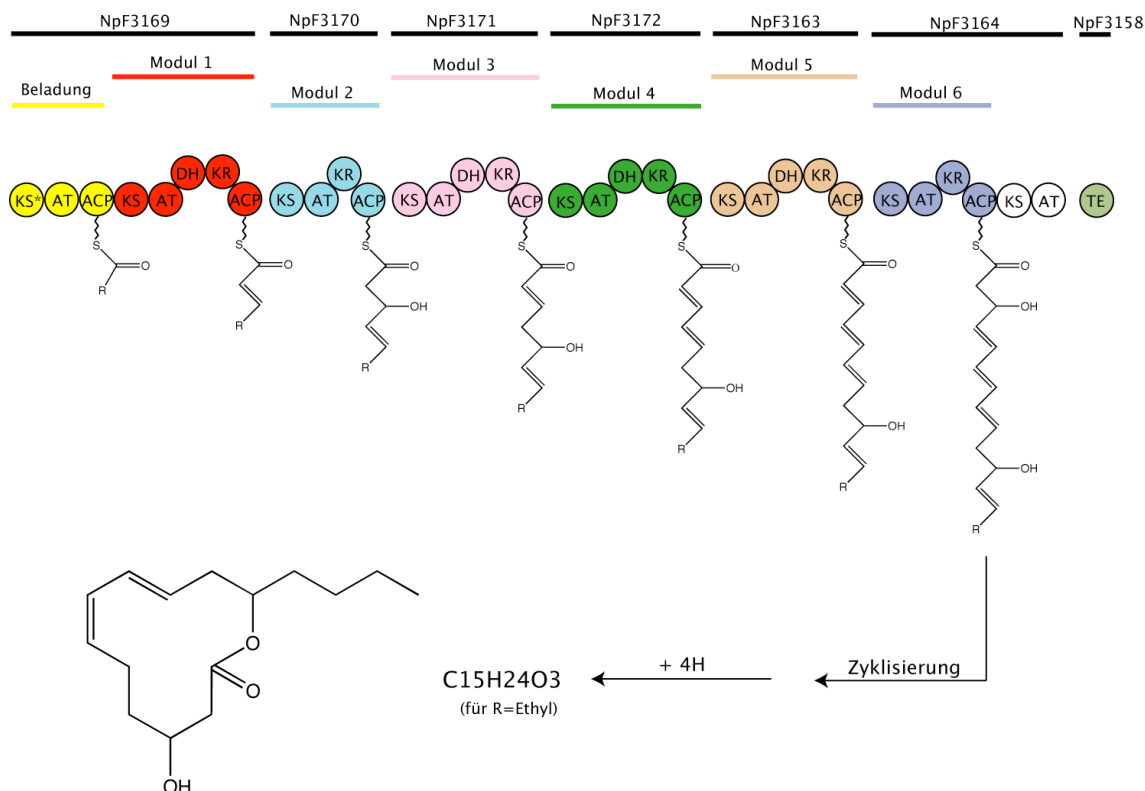


Abb. 55: Vorschlag eines Biosynthese-Schemas für den von Cluster VII produzierten Metaboliten.

Das Schema ist spekulativ und beruht auf den im Text genannten Annahmen über die Inaktivität bestimmter Module. Die verschiedenen Module sind farblich gegeneinander abgehoben. R bezeichnet einen allgemeinen Rest.

4.4 Evolution von bakteriellen PKS-I und NRPS im Kontext der Generierung metabolischer Diversität

Abschließend sollen die über die Evolution von PKS und NRPS in Bakterien gewonnenen Erkenntnisse in den allgemeinen Zusammenhang der Evolution des Sekundärmetabolismus eingeordnet werden. Wie in der Einleitung (siehe Abschnitt 1.1.1) bereits dargestellt, handelt es sich bei der begrifflichen Abgrenzung des Terminus Sekundärmetabolismus um eine *ad hoc*-Definition, die sich im wesentlichen auf die nicht generelle Verbreitung bestimmter Metabolite bezieht, nicht jedoch auf den Aspekt der biologischen Funktion. Obwohl die Begriffe Primärstoffwechsel und Sekundärstoffwechsel allgemein in Gebrauch sind, ist die Abgrenzung nicht immer klar, wie daran zu erkennen ist, dass sich eine Reihe von Verbindungen nicht ohne weiteres einer der beiden Kategorien zuordnen lassen. Beispielsweise sind Lipide und Polysaccharide von essenzieller Bedeutung für alle Organismen, einige Vertreter dieser Substanzklassen spielen jedoch für den produzierenden Organismus eine besondere Rolle, die sie in anderen Lebewesen nicht haben (Firn und Jones 2000). Diese Funktionen von Sekundärmetaboliten beruhen auf der häufig

anzutreffenden hochwirksamen biomolekularen Aktivität, die eine spezifische Wechselwirkung mit bestimmten Zielmolekülen ermöglicht. So sind Sekundärmetabolite Bestandteile einer Art von „biologischer Kriegsführung“ gegen Eindringlinge, Fraßfeinde und Parasiten oder übernehmen wichtige physiologische Funktionen (Vining 1992). Aus evolutionärer Sicht lautet die zentrale Frage: Welche Mechanismen führen zu dieser extremen Strukturdiversität? Die Antwort auf diese Frage ist nicht allein von eigenständig grundsätzlichem Interesse, sondern könnte auch in Zukunft die Screening-Protokolle pharmazeutischer Unternehmen verbessern helfen und hilfreiche Rückschlüsse auf Design biotechnologischer Ansätze zur künstlichen Generierung von möglichen Wirkstoffmolekülen zulassen.

4.4.1 Generierung von Strukturdiversität

Die im Sekundärstoffwechsel vorkommenden Enzyme lassen sich in zumindest zwei große Gruppen unterteilen: Zum einen existieren spezielle Biosynthese-Systeme wie PKS und NRPS, welche die Basis der vorliegenden Arbeit bilden. Diese Systeme findet man ausschließlich im Sekundärstoffwechsel. Zum anderen kennt man auch zahlreiche „klassische“ Stoffwechsel-Ketten aus Einzelenzymen, wie sie typisch für den Primärstoffwechsel sind.

Ein einzigartiges enzymatisches Prinzip zur Strukturdiversifizierung stellen die modularen PKS des Typs I dar. Die Möglichkeit, durch das jeweilige Modul-Design leichte Modifikationen in der Polyketid-Kette hervorzurufen, führt zu einer enormen Vielfalt an unterschiedlichen Strukturen, obwohl nur vier unterschiedliche Typen von Modulen zusammengeschaltet werden (siehe **Abb. 5**). Die Strukturunterschiede zwischen den diversen Polyketid-Verbindungen beruhen dabei auf der quasi freien Kombinierbarkeit der folgenden „Variablen“:

- (1) Auswahl des einzubauenden Bausteins, d. h. während der Kettenverlängerung im wesentlichen die Entscheidung zwischen Malonyl-CoA, und Methylmalonyl-CoA, aber auch anderen α -Alkylmalonyl-CoA-Varianten, wobei letzter insbesondere in großer Vielfalt bei der Selektion des Substrates durch das Beladungsmodul vorkommen (siehe **Tab. 1**),
- (2) Bildung unterschiedlicher Stereoisomere, falls wie etwa im Fall von Methylmalonyl-CoA ein chirales Zentrum vorhanden ist,
- (3) Bildung unterschiedlicher Oxidationsstufen durch die optionale Aktivität der Domärentypen KR, DH und ER (siehe **Abb. 5**),
- (4) Bildung unterschiedlicher Stereoisomere, falls nur eine KR-Domäne aktiv ist und keine Dehydratase- und Enoylreductase-Reaktionen erfolgen (siehe **Abb. 5**).

Berücksichtigt man diese Möglichkeiten der Reaktionsvariierung, so lässt sich für ein PKS-System aus sechs Elongationsmodulen berechnen, dass es mehr als 100.000 unterschiedliche Verbindungen zu produzieren vermag (Gonzalez-Lergier, Broadbelt und Hatzimanikatis 2005). Wie in den Abschnitten 3.3 und 4.2.3 dargestellt und diskutiert wurde, konnte im Rahmen dieser Arbeit gezeigt werden, dass das Actinobakterium *Streptomyces avermitilis* bei der Entwicklung von entsprechenden Genclustern

ausgiebigen Gebrauch von der Neukombinationen von Modulen macht, also gewissermaßen „natürliche Biokombinatorik“ betreibt und damit verblüffende Ähnlichkeiten zu diversen experimentellen Laboransätzen offenbart. Besonders betont werden muss in diesem Zusammenhang, dass der Aufbau selbst von sehr großen Genclustern dieser Art wahrscheinlich in relativ kurzer Zeit möglich ist. Diese Schlussfolgerung lässt sich jedenfalls aus der Struktur des Clusters für die Mycolacton-Biosynthese in *Mycobacterium ulcerans* ziehen. Die KS-Domänen dieser Synthase weisen auf Aminosäure-Ebene eine Identität von mindestens 97 % auf, was auf nicht lange zurück liegende Duplikationsereignisse schließen lässt (Stinear *et al.* 2004).

Modular gebaute Biosynthese-Systeme können als molekulare Maschinen aufgefasst werden, die von der Natur erfunden wurden, um ein Maximum an chemischer Strukturvariabilität auf „Gen-sparende“ Weise generieren zu können; und Strukturvariabilität ist die Grundvoraussetzung für die Herausbildung spezifischer biomolekularer Wechselwirkungen. Darauf wird weiter unten nochmals vertieft eingegangen. Kleine Moleküle, wie typische Sekundärmetabolite sie darstellen, mit Hilfe riesiger Enzymkomplexe zu synthetisieren, macht nur vor dem Hintergrund der darin integrierten Flexibilität des Prozesses Sinn. Tatsächlich wird dieser Prozess häufig bildlich sehr treffend als Fließband-Produktion beschrieben. Hieraus wird auch verständlich, warum diese Art von Biosynthese-Systemen nur im Bereich des Sekundärstoffwechsels auftritt. Sie produzieren zu einem bestimmten Zeitpunkt eine bestimmte Verbindung, tragen aber außerdem beständig das Potenzial in sich, diese Verbindung durch einfache Rekombinationsereignisse auf genetischer Ebene zu modifizieren.

Neben diesen speziellen Systemen kommen auch viele aus einzelnen Enzymen aufgebaute Stoffwechselketten im Sekundärmetabolismus vor. So etwa in der Biosynthese der Terpene, der Alkaloide und der zahlreichen Phenylpropan-Derivate. Hier stellt sich die Frage, inwiefern besonder Eigenschaften der beteiligten Enzyme zur Strukturvielfalt beitragen. Es wurde vorgeschlagen, dass diese Enzyme – zumindest in manchen Fällen – eine gering ausgeprägte Substratspezifität aufweisen und darüber hinaus nicht in eindimensionalen Ketten, sondern verzweigt oder als Matrix-System organisiert sind (Firn und Jones 2000). Im Zusammenhang der von ihnen vorgeschlagenen so genannten „Screnning-Hypothese“ haben Firn und Jones die Literatur nach Hinweisen auf solche Eigenschaften durchsucht und eine beeindruckende Zahl an Beispielen zusammengetragen (Firn und Jones 2003). So konnte gezeigt werden, dass eine Mutation in einem einzigen Enzym in *Mentha gracilis* (schottische Grüne Minze) zur Produktion einer Reihe von neuen *p*-Methanmonoterpenen führte. Diese Mutation bewirkte eine Oxygenierung in der Position C3 anstatt der üblichen Position am C6, ergab jedoch nicht nur eine neue Verbindung, sondern eine ganze Serie, da die auf den Oxygenierungsschritt folgenden Enzyme der Kette in der Lage sind, unterschiedliche Substrate zu verarbeiten (Croteau *et al.* 1991). Ein Fall einer Gitter-artig organisierten Biosynthese wurde in *Agrobacterium aurantiacum* in der Astaxanthin-Biosynthese gefunden. Das Endprodukt Astaxanthin kann, ausgehend von β -Caroten, über acht verschiedene Zwischenprodukte erhalten werden, wobei an diesen Umsetzungen nur zwei Enzyme beteiligt sind, eine Hydroxylase und eine Ketolase (Misawa *et al.* 1995). Für PKS des Typs III in Pflanzen konnte gezeigt werden, dass ihre Evolution dem Modell der Genduplikationen mit anschließenden Punktmutationen im aktiven Zentrum folgt, was infolge unterschiedlicher Zyklisierungsmechanismen zu verschiedenen Produkten führt. Auf diese Weise haben sich Stilben-Synthasen mehrmals unabhängig voneinander aus Chalcon-Synthasen entwickelt (Tropf *et al.* 1994; Durbin *et al.* 1995; Helariutta *et al.* 1996). Somit kann konstatiert werden:

Eine Reihe von Hinweisen spricht dafür, dass Enzyme, die eine Rolle im Sekundärstoffwechsel spielen, auf besondere Weise organisiert sind oder – gerade im Hinblick auf ihre Substratspezifität – besondere Eigenschaften aufweisen, so dass sie Produkte von breiter chemischer Diversität herstellen können. Allerdings ist dies kein durchgehendes Merkmal, das auf *alle* Enzyme dieses Stoffwechselbereichs zutrifft. Vielmehr sind ebenso zahlreiche Beispiele bekannt, die sich durch sehr stark ausgeprägte Substratspezifitäten auszeichnen, etwa in den Biosynthese-Wegen der Phenylpropan-Derivate und der Alkaloide (Zenk 1991).

4.4.2 Hypothesen zur funktionellen Bedeutung von Sekundärmetaboliten

Die natürliche Selektion, der alle Organismen unterliegen, fördert vorteilhafte Merkmale, nämlich solche, welche die Fitness des Organismus erhöhen. Die Frage der Selektion ist daher direkt mit der Frage nach der biologisch funktionellen Bedeutung von Verbindungen des Sekundärstoffwechsels verbunden. Die enorme Diversität dieser Metabolite aus Pflanzen und Mikroorganismen hat immer wieder zu Spekulationen und Debatten hinsichtlich ihrer biologischen Rolle herausgefordert. Eine vor allem in der älteren Literatur häufig geäußerte Hypothese besagt, Sekundärmetabolite seien als Abfall- bzw. Detoxifikationsprodukte von Organismen anzusehen (Malik 1980). Weiterhin wird des öfteren die Meinung vertreten, dass diese Substanzen gewissermaßen ein Arsenal „chemischen Kriegsführung“ zwischen Lebewesen darstellen (Baba und Schneewind 1998). Gegner dieser Ansicht brachten dagegen die Idee einer „nicht funktionellen Stoffvielfalt“ ins Spiel, die als Basis für strukturelle Neuerfindungen dient (Zähler und Anke 1983). Außerdem wurde vorgeschlagen, diese Moleküle könnten die Überreste früher funktioneller Metabolite sein, die jetzt jedoch keine Funktion mehr aufweisen (Davies 1990).

Die genannten Meinungen sind nicht als abschließende Aufzählung gedacht, sondern sollen einen Eindruck davon vermitteln, welche – häufig durchaus einflussreichen – Vorschläge zu diesem Thema im Verlauf der letzten hundert Jahre gemacht wurden. Einige dieser Vorschläge sind sicherlich nicht geeignet, den gesamten Bereich des Sekundärstoffwechsels zu deuten, andere sind mittlerweile durch experimentelle Daten eindeutig widerlegt.

Die Ansicht, Sekundärmetabolite seien einfach Abfallprodukte des Primärstoffwechsels spiegelt deutlich wider, dass in der ersten Hälfte des 20. Jahrhunderts praktisch nichts über die biologischen Funktionen dieser Verbindungen bekannt war. Das grundsätzliche Problem dieser Hypothese liegt in ihrer Begrenztheit; sie gibt keinerlei Aufschlüsse darüber, warum so viele, strukturell stark unterschiedliche Sekundärmetabolite vorkommen. Im übrigen läuft sie den physiologischen Verhältnissen bei Pflanzen und Mikroorganismen, also den Hauptproduzenten dieser Verbindungen, zuwider, denn gerade diese Organismen würden eine solche Art von Detoxifizierungssystem nicht benötigen: Pflanzen sind in der Lage, ihren metabolischen „Input“ genau zu kontrollieren, und von Mikroorganismen wäre zu erwarten, dass etwaige Abfallstoffe nach außen transportiert und an die Umgebung abgegeben werden, anstatt im Laufe der Evolution eine große Zahl an Energie-aufwändigen Reaktionen zu entwickeln. Falls Stoffwechselwege zum Zweck der Detoxifizierung zu einem frühen Zeitpunkt der Evolution eine essenzielle Rolle gespielt hätten, so wäre eine Konservierung der notwendigen Reaktionen in verschiedenen Organismengruppen das zu erwartende Ergebnis, nicht jedoch eine überbordende Vielfalt.

Schließlich hat sich der Kenntnisstand hinsichtlich der physiologischen Funktionen von Sekundärmetaboliten deutlich verbessert. Es gibt inzwischen viele Beispiele von Verbindungen, denen eine wichtige Funktion im Rahmen ökologischer Wechselwirkungen zugeordnet werden konnte (siehe hierzu z. B. (Berenbaum und Robinson 2003; Hartmann 2004)). Fasst man diese Überlegungen zusammen, so kann zwar nicht ausgeschlossen werden, dass *einige* Verbindungen des Sekundärstoffwechsels im Zusammenhang mit Abfall- oder Detoxifizierungsprodukten stehen oder ursprünglich standen, allerdings ist dieser Erklärungsansatz weit von einem allgemeinen Prinzip entfernt.

Sekundärmetabolite als Spuren der präbiotischen Evolution zu interpretieren, war ein – durchaus provokativer – Vorschlag von Davies (Davies 1990). Gemäß dieser Spekulation könnten sie ursprünglich zur Stabilisierung von Biopolymeren in einer aktiven Konformation beigetragen oder sogar als eine Art primitiver Katalysator gewirkt haben. Beispielsweise könnten solche Antibiotika, welche die diversen Abläufe am Ribosom inhibieren, Bestandteile eines frühen Translations-ähnlichen Systems gewesen sein. Auch diese Hypothese erscheint aus mehreren Gründen unwahrscheinlich. Zunächst ist festzustellen, dass Biopolymere wie Nucleinsäuren und Proteine, abgesehen von Wasser, typischerweise keine kleinen Moleküle zur Stabilisierung benötigen. Zweitens scheint es nach heutigem Kenntnisstand für die in Frage stehenden präbiotischen Reaktionen keine treibenden Kräfte gegeben zu haben, die *auf selektive Weise* zu bestimmten Verbindungen geführt hätten. Zum dritten kann derselbe allgemeine Einwand wie im Fall der „Abfall-Hypothese“ erhoben werden. Weshalb nämlich auf diese Weise eine so große chemisch-strukturelle Vielfalt entstanden sein sollte, lässt sich auf diese Weise nicht erklären.

In den letzten Jahren hat sich eine andere Auffassung zunehmend durchgesetzt: Sekundärmetabolite werden nunmehr als Träger der Kommunikation in komplexen biologischen Vergesellschaftungen erkannt (Davies 2006). Ein wichtiges Argument, um diese Sichtweise zu stützen, ist der Hinweis, dass Substanzen, denen eine antibiotische Wirkung zueigen ist, unter den realen physiologischen und ökologischen Bedingungen des produzierenden Organismus nicht die notwendigen Konzentrationen erreichen. Andererseits ist jedoch bekannt, dass auch Konzentrationen unterhalb der für die inhibitorische Wirkung kritischen Konzentration zu weitreichenden Änderungen der Transkription führen können, und zwar sowohl beim produzierenden Organismus als auch bei den natürlicherweise im Habitat vorkommenden Nachbarn. Eine Reihe neuerer Publikationen legt eine Rolle von Sekundärmetaboliten in den Signalwegen der zellulären Kommunikation nahe. Es konnte z. B. gezeigt werden, dass Antibiotika wie Rifampicin und Erythromycin in geringen, subinhibitorischen Konzentrationen die globalen Transmissionsmuster in Bakterien beeinflussen, was mit Hilfe von Promotor-*lux*-Konstrukten in *Salmonella typhimurium* gemessen wurde (Goh *et al.* 2002). In dem Myxobakterium *Myxococcus xanthus* haben die so genannten DK-Xanthene großen Einfluss auf die Ausbildung des Myxobakterientypischen Fruchtkörpers und auf die Entwicklungsfähigkeit der Sporen (Meiser, Bode und Müller 2006). Ein weiteres Beispiel für den Einfluss von Sekundärmetaboliten auf zelluläre Interaktionen stellt das cyanobakterielle Toxin Microcystin dar. Diese Verbindung bewirkt Veränderungen in der Expression und der Oligomerisierung eines Lectins auf der Zelloberfläche von *Microcystis aeruginosa* und beeinflusst somit die Zell-Zell-Erkennung und den Zusammenhalt der einzelnen Zellen in Zellkolonien (Kehr *et al.* 2006).

Auf dem Gebiet des pflanzlichen Sekundärstoffwechsels war die Idee der „chemischen Kriegsführung“ mit Hilfe bestimmter Metabolite schon länger bekannt. Ebenso waren die wichtigen

Beziehungen zwischen phytochemischer Diversität und ökologischen Interaktionen bereits seit den 1960er Jahren in das Blickfeld der Forscher gerückt. Die Hypothese entstand durch die Verknüpfung zweier grundlegender Beobachtungen: die enorme Diversität pflanzlicher Sekundärmetabolite und die gleichsam ausgeprägte Vielfalt der ökologischen Beziehungen zwischen Pflanzen und Insekten (Fraenkel 1959). Diese Sichtweise konnte in der Folgezeit durch viele Studien weiter gestützt werden, die sich der Koevolution innerhalb ökologischer Systeme annahmen. Verbindungen des pflanzlichen Sekundärstoffwechsels können beispielsweise den Signalaustausch bei symbiotischen Beziehungen modulieren, wie für die Flavon-Verbindung Luteolin gezeigt werden konnte. Luteolin induziert die Expression des *nod*-Gens in den Knöllchen-bildenden Rhizobakterien (Peters, Frost und Long 1986). Pflanzen haben im Verlauf der Evolution auch chemische Reaktionswege gegen Insekten-Fraßfeinde herausgebildet, wobei sich direkte und indirekte Strategien unterscheiden lassen (Kessler und Baldwin 2002). Beispiele für eine direkte chemische Abwehrstrategie sind Nikotin in *Nicotiana attenuata* (Winz und Baldwin 2001) und die verschiedenen in Angiospermen zu findenden Pyrrolizidin-Alkaloide (Hartmann 2004). Zu den indirekten Strategien gehört das Anlocken natürlicher Feinde der Fraßinsekten mit Hilfe leicht flüchtiger Substanzen, die im Englischen unter der Bezeichnung „volatile organic substances“ (häufig abgekürzt als VOC) bekannt sind. Gut untersucht ist etwa die Freisetzung von Terpenen und Indol durch Maispflanzen. Induziert wird die Freisetzung durch ein Fettsäure-Derivat, das *N*-(17-Hydroxylinolenoyl)-L-Glutamin. Diese Verbindung ist ein typischer Bestandteil des Speichels eines Fraßfeindes, nämlich der Raupen des Großen Kohlweißlings (*Pieris brassicae*). Kommt die Maispflanze durch Fraßwunden in Kontakt mit dem Fettsäure-Derivat, so lockt sie mit Hilfe des in die Atmosphäre abgegebenen Terpen-Indol-Gemisches Schlupfwespen (*Cortesia marginiventris*) an, die ihrerseits natürlicher Fraßfeind der Raupen sind (Frey *et al.* 2000). Darüber hinaus ließ sich auch zeigen, dass Pflanzen auf das Ablegen von Insekteneiern prophylaktisch reagieren können, indem sie Lockstoffe in die Luft abgeben, die entsprechende Fraßinsekten anziehen (Hilker *et al.* 2002). Auch diese Stoffe müssen aufgrund ihrer Verteilung als typische Sekundärmetabolite angesprochen werden. Abschließend sei noch darauf hingewiesen, dass Sekundärmetabolite in Pflanzen Hormonfunktionen besitzen können, wie am Beispiel der Brassinosteroide, die auch unter der Bezeichnung Brassine bekannt sind, zu sehen ist (Szekeres *et al.* 1996).

Anhand der wenigen hier angeführten Beispiele wird erkennbar, dass mittlerweile in vielen Fällen die physiologische Funktion von Produkten des Sekundärstoffwechsels außer Frage steht. Die Heranbildung dieser Verbindungen mit ihren sehr spezifischen molekularen Wechselwirkungen ist nur durch Adaptationsprozesse auf der Basis natürlicher Selektion erklärbar. Es bleibt allerdings die Frage, ob man schlussfolgern kann, *alle* Sekundärmetabolite hätten eine – klar umrissene – biologische Funktion. Dies führt wiederum auf die Frage nach der Bedeutung von Selektionsprozessen in der Evolution des Sekundärmetabolismus im Allgemeinen und begründet eine Diskussion, die sich in den letzten Jahren zugespitzt hat und aus der Sicht zweier wissenschaftlicher „Lager“ sehr konträr geführt wird.

4.4.3 Welche Rolle spielen Selektion und Adaptation in der Evolution des Sekundärstoffwechsels?

Die Meinungsverschiedenheiten in dieser Frage finden ihren Ausdruck unter anderem in einem 1996 veröffentlichten Übersichtsartikel mit dem Titel: „Phytochemical diversity: adaptation or random variation?“ (Berenbaum und Zangerl 1996). Es handelt sich dabei insofern um ein sehr bezeichnendes Beispiel der grundsätzlichen Diskussion, als der Eindruck erweckt wird, dass für die Klärung der Diversität des Sekundärstoffwechsels eine Entscheidung zwischen diesen beiden Alternativen getroffen werden müsste. Stillschweigend wird dabei angenommen, die Alternativen würden sich gegenseitig ausschließen. In diesem Zusammenhang muss an einige prinzipielle Erkenntnisse der Evolutionsbiologie erinnert werden (siehe hierzu z. B. (Ridley 2004)). Natürliche Selektion ist die einzige Erklärung für Adaptationsprozesse, und natürliche Selektion kann nur dann auf eine Population einwirken, wenn es Variation zwischen den Angehörigen der Population gibt. Diese Variation ist selbstverständlich zufällig im Hinblick auf die Richtung der sich vollziehenden Adaptation. Mit anderen Worten: Zufällige (d. h. ungerichtete) Variation ist Grundlage der natürlichen Selektion, und diese wiederum ist Grundlage der Adaptation. Deshalb liegt kein erkennbarer Sinn darin, Variation und Adaptation in ein alternatives Verhältnis zu bringen.

Wendet man dies auf das Problem der Evolution des Sekundärmetabolismus an, so gilt: Je mehr Variation molekularer Strukturen ein Organismus hervorzubringen in der Lage ist, desto höher ist die Wahrscheinlichkeit, dass es ihm gelingt, einen geeigneten „Startpunkt“ für sich anschließende Adaptationsprozesse zu finden. Adaptation führt dann zu Verbindungen, die in spezifische physiologische Abläufe eingreifen und etwa mit extremer Affinität an bestimmte Rezeptoren binden, wie man es z. B. bei DNA-bindenden Antibiotika wie dem Bleomycin beobachten kann (Williams *et al.* 1989). Eine derartige Bindungsgenauigkeit erklärt sich am besten folgendermaßen: Zunächst gab es ein mehr oder weniger gut bindendes Molekül, dass mit dieser Anfangsstruktur völlig zufällig entstand. Im Verlauf einer andauernden ökologischen Interaktion durchlief das Molekül bzw. das synthetisierende Enzymsystem einen Prozess der Feinabstimmung. Ohne diese Wechselwirkung mit der Umgebung wäre diese Ausgangsstruktur – ebenso wie viele andere vor ihr – aus dem Stoffwechsel entfernt worden. Dieses Konzept impliziert allerdings zwangsläufig, dass ein gewisser Anteil dieser Verbindungen, die unter dem Begriff Sekundärmetabolismus zusammengefasst werden, zu einem bestimmten Zeitpunkt keine oder zumindest praktisch keine spezifische Funktion besitzt.

Hiervon ausgehend schlugen Firn und Jones Anfang der 1990er Jahre ihre so genannte „Screening-Hypothese“ vor (Jones und Firn 1991), die sie in den folgenden 15 Jahren beständig verfeinerten und ausbauten (Firn und Jones 2000; Firn und Jones 2003). Die „Screening-Hypothese“ fußt auf der Erkenntnis, dass für ein beliebiges gegebenes Molekül die Wahrscheinlichkeit, eine starke biologische Aktivität zu besitzen, sehr gering ist. Daher sind Organismen darauf angewiesen, so viel chemische Variabilität wie möglich zu erzeugen. Denn dadurch steigt die Wahrscheinlichkeit, ein Molekül zu generieren, das zu einer bestimmten, sehr spezifischen Funktion passt. Von diesem Standpunkt aus lässt sich der Sekundärmetabolismus als anhaltendes Ergebnis des Zusammenspiels aus Strukturgenerierung und dem „Screening“ dieser Substanzbibliotheken nach Funktionen im physiologischen/ökologischen Zusammenhang deuten.

Allerdings gibt es in der Wissenschaft deutliche Widerstände gegen diese Auffassung, und es wird heftig über die Frage diskutiert, ob wir unseren Kenntnisstand zur biologischen Funktion von Sekundärmetaboliten als Kenntnislücken verstehen sollen oder ob tatsächlich – möglicherweise sehr viele – Verbindungen ohne konkrete Funktion existieren. Die neuesten Beiträge zu dieser Debatte beschäftigen sich mit leicht flüchtigen Metaboliten aus Pflanzen (Owen und Penuelas 2005; Firn und Jones 2006a; Firn und Jones 2006b; Pichersky, Sharkey und Gershenzon 2006). Ausgehend von einem Übersichtsartikel über die Emissionsmuster flüchtiger Isoprenoide und ihrer möglichen Interpretation, entwickelte sich schließlich ein Disput über die Richtigkeit der „Screening-Hypothese“, deren Kern eben in der Annahme besteht, dass viele Sekundärmetabolite keinerlei Funktion aufweisen. Dagegen wird häufig der Einwand erhoben, die Existenz von Verbindungen ohne Funktion sei prinzipiell inkonsistent mit der Evolutionstheorie. Dahinter steht die Überlegung, dass die Biosynthese der Sekundärmetabolite „Fitness-Kosten“ hervorrufe und deshalb solche Genotypen, die funktionslose und somit nutzlose Metaboliten hervorbringen, ausselektiert würden (Berenbaum und Zangerl 1996). Da die Selektion aufgrund der aktuellen Fitness eines Organismus geschieht, nicht aber im Hinblick auf mögliche zukünftige Potenziale, sollte es keine oder so gut wie keine funktionslosen Verbindungen geben. Die Konsequenz hieraus wäre, dass der ganz überwiegenden Mehrzahl an Sekundärmetaboliten eine physiologische Funktion unterstellt werden müsste, die dem produzierenden Organismus einen konkreten Selektionsvorteil bringt.

Sicherlich haben weit mehr Verbindungen des Sekundärmetabolismus eine konkrete biologische Funktion im Umweltzusammenhang des jeweiligen Organismus als bisher bekannt. Das beschriebene Argument der „Fitness-Kosten“ ist jedoch in dieser allgemeinen Form nicht völlig überzeugend. So müsste im Rahmen dieser Argumentation auch die Zeit berücksichtigt werden, die notwendig ist, um gegen einen Sekundärmetabolit-produzierenden Genotyp zu selektieren. Es fällt auf, dass sich die Diskussion auf qualitative Argumente beschränkt, während quantitative Studien fehlen. Denkt man ausschließlich in der Kategorie von „Fitness-Kosten“, ohne dabei die zeitliche Dimension und eine Abschätzung dieser Kosten mit einzubeziehen, so würde dies letztlich den Schluss nahelegen, dass keinerlei genetische Verbesserung möglich wäre. Denn alle Genduplikationen führen zeitweise zu erhöhten „Fitness-Kosten“, da es in jedem Fall einer gewissen Zeitspanne bedarf, bis sich neue, gewinnbringende Funktionen herausgebildet haben.

Die weitere Erforschung von Enzymen des Sekundärmetabolismus gerade im Hinblick auf die phylogenetischen Verhältnisse und die physiologischen Funktionen der produzierten Verbindungen wird es ermöglichen, gut begründete und quantitative Modelle der Evolution metabolischer Diversität zu entwickeln.

5 Literatur

- Abascal, F., R. Zardoya und D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**:2104-2105.
- Ackerley, D. F., T. T. Caradoc-Davies und I. L. Lamont. 2003. Substrate specificity of the nonribosomal peptide synthetase PvdD from *Pseudomonas aeruginosa*. *J Bacteriol* **185**:2848-2855.
- Adams, D. G. 2000. Heterocyst formation in cyanobacteria. *Curr Opin Microbiol* **3**:618-624.
- Adl, S. M., A. G. Simpson, M. A. Farmer, R. A. Andersen, O. R. Anderson, J. R. Barta, S. S. Bowser, G. Brugerolle, R. A. Fensome, S. Fredericq, T. Y. James, S. Karpov, P. Kugrens, J. Krug, C. E. Lane, L. A. Lewis, J. Lodge, D. H. Lynn, D. G. Mann, R. M. McCourt, L. Mendoza, O. Moestrup, S. E. Mozley-Standridge, T. A. Nerad, C. A. Shearer, A. V. Smirnov, F. W. Spiegel und M. F. Taylor. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol* **52**:399-451.
- Allen, M. B. und D. I. Arnon. 1955. Studies on Nitrogen-Fixing Blue-Green Algae. I. Growth and Nitrogen Fixation by *Anabaena Cylindrica* Lemm. *Plant Physiol* **30**:366-372.
- Alonso, W. R. und R. Croteau. 1993. Prenyltransferases and cyclases. Pp. 239-260 *in* P. J. Lea, ed. *Enzymes of secondary metabolism*. Academic Press, London.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers und D. J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* **215**:403-410.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller und D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**:3389-3402.
- Amy, C. M., B. Williams-Ahlf, J. Naggert und S. Smith. 1992. Intron-exon organization of the gene for the multifunctional animal fatty acid synthase. *Proc Natl Acad Sci U S A* **89**:1105-1108.
- Aparicio, J. F., I. Molnar, T. Schwecke, A. König, S. F. Haydock, L. E. Khaw, J. Staunton und P. F. Leadlay. 1996. Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**:9-16.
- Arnez, J. G. und D. Moras. 1997. Structural and functional considerations of the aminoacylation reaction. *Trends Biochem Sci* **22**:211-216.
- August, P. R., L. Tang, Y. J. Yoon, S. Ning, R. Müller, T. W. Yu, M. Taylor, D. Hoffmann, C. G. Kim, X. Zhang, C. R. Hutchinson und H. G. Floss. 1998. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699. *Chem Biol* **5**:69-79.
- Austin, M. B. und J. P. Noel. 2003. The chalcone synthase superfamily of type III polyketide synthases. *Nat Prod Rep* **20**:79-110.
- Azad, A. K., T. D. Sirakova, N. D. Fernandes und P. E. Kolattukudy. 1997. Gene knockout reveals a novel gene cluster for the synthesis of a class of cell wall lipids unique to pathogenic mycobacteria. *J Biol Chem* **272**:16741-16745.
- Baba, T. und O. Schneewind. 1998. Instruments of microbial warfare: bacteriocin synthesis, toxicity and immunity. *Trends Microbiol* **6**:66-71.
- Baldauf, S. L. und J. D. Palmer. 1993. Animals and fungi are each other's closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci U S A* **90**:11558-11562.
- Baldauf, S. L., A. J. Roger, I. Wenk-Siefert und W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**:972-977.
- Baldock, C., J. B. Rafferty, S. E. Sedelnikova, S. Bithell, A. R. Stuitje, A. R. Slabas und D. W. Rice. 1996. Crystallization of *Escherichia coli* enoyl reductase and its complex with diazaborine. *Acta Crystallogr D Biol Crystallogr* **52**:1181-1184.
- Bangera, M. G. und L. S. Thomashow. 1999. Identification and characterization of a gene cluster for synthesis of the polyketide antibiotic 2,4-diacetylphloroglucinol from *Pseudomonas fluorescens* Q2-87. *J Bacteriol* **181**:3155-3163.
- Baptiste, E. und C. Brochier. 2004. On the conceptual difficulties in rooting the tree of life. *Trends Microbiol* **12**:9-13.
- Barry, C. E., 3rd, R. E. Lee, K. Mdluli, A. E. Sampson, B. G. Schroeder, R. A. Slayden und Y. Yuan. 1998. Mycolic acids: structure, biosynthesis and physiological functions. *Prog Lipid Res* **37**:143-179.

- Battistuzzi, F. U., A. Feijao und S. B. Hedges. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy und the colonization of land. *BMC Evol Biol* **4**:44.
- Beck, B. J., Y. J. Yoon, K. A. Reynolds und D. H. Sherman. 2002. The hidden steps of domain skipping: macrolactone ring size determination in the pikromycin modular polyketide synthase. *Chem Biol* **9**:575-583.
- Bedford, D., J. R. Jacobsen, G. Luo, D. E. Cane und C. Khosla. 1996. A functional chimeric modular polyketide synthase generated via domain replacement. *Chem Biol* **3**:827-831.
- Belshaw, P. J., C. T. Walsh und T. Stachelhaus. 1999. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**:486-489.
- Bender, C. L., F. Alarcon-Chaidez und D. C. Gross. 1999. *Pseudomonas syringae* phytotoxins: mode of action, regulation und biosynthesis by peptide and polyketide synthetases. *Microbiol Mol Biol Rev* **63**:266-292.
- Bentley, S. D., K. F. Chater, A. M. Cerdeno-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C. H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabinowitsch, M. A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill und D. A. Hopwood. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**:141-147.
- Berenbaum, M. R. und G. E. Robinson. 2003. Chemical communication in a post-genomic world. *Proc Natl Acad Sci U S A* **100 Suppl 2**:14513.
- Berenbaum, M. R. und A. R. Zangerl. 1996. Phytochemical diversity: adaptation or random variation? Pp. 1-24 in J. T. Romeo, J. A. Saunders und P. Barbosa, eds. *Phytochemical diversity and redundancy in ecological interactions*. Plenum Press, New York.
- Bergmann, S., J. Schumann, K. Scherlach, C. Lange, A. A. Brakhage und C. Hertweck. 2007. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nat Chem Biol* **3**:213-217.
- Bevitt, D. J., J. Cortes, S. F. Haydock und P. F. Leadlay. 1992. 6-Deoxyerythronolide-B synthase 2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. *Eur J Biochem* **204**:39-49.
- Biondi, N., R. Piccardi, M. C. Margheri, L. Rodolfi, G. D. Smith und M. R. Tredici. 2004. Evaluation of *Nostoc* strain ATCC 53789 as a potential source of natural pesticides. *Appl Environ Microbiol* **70**:3313-3320.
- Birch, A. J. 1967. Biosynthesis of polyketides and related compounds. *Science* **156**:202-206.
- Birky, C. W., Jr. und R. V. Skavaril. 1976. Maintenance of genetic homogeneity in systems with multiple genomes. *Genet Res* **27**:249-265.
- Bisang, C., P. F. Long, J. Cortes, J. Westcott, J. Crosby, A. L. Matharu, R. J. Cox, T. J. Simpson, J. Staunton und P. F. Leadlay. 1999. A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**:502-505.
- Black, T. A., Y. Cai und C. P. Wolk. 1993. Spatial expression and autoregulation of *hetR*, a gene involved in the control of heterocyst development in *Anabaena*. *Mol Microbiol* **9**:77-84.
- Blokhin, A. V., H. D. Yoo, R. S. Gerald, D. G. Nagle, W. H. Gerwick und E. Hamel. 1995. Characterization of the interaction of the marine cyanobacterial natural product curacin A with the colchicine site of tubulin and initial structure-activity studies with analogues. *Mol Pharmacol* **48**:523-531.
- Bode, H. B., B. Bethe, R. Hofs und A. Zeeck. 2002. Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem* **3**:619-627.
- Bode, H. B. und R. Müller. 2005. The impact of bacterial genomics on natural product research. *Angew Chem Int Ed Engl* **44**:6828-6846.
- Bode, H. B. und R. Müller. 2006. Analysis of myxobacterial secondary metabolism goes molecular. *J Ind Microbiol Biotechnol* **33**:577-588.
- Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**:248-254.
- Brattsten, L. B. 1979. Biochemical defense mechanisms in herbivores against plant allochemicals. Pp. 199-270 in G. A. Rosenthal und D. H. Janzen, eds. *Herbivores, their interaction with secondary plant metabolites*. Academic Press, New York.
- Brennan, P. J. und H. Nikaido. 1995. The envelope of mycobacteria. *Annu Rev Biochem* **64**:29-63.

- Brochier, C., E. Baptiste, D. Moreira und H. Philippe. 2002. Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet* **18**:1-5.
- Brocks, J. J., G. A. Logan, R. Buick und R. E. Summons. 1999. Archean molecular fossils and the early rise of eukaryotes. *Science* **285**:1033-1036.
- Caffrey, P. 2003. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *Chembiochem* **4**:654-657.
- Campbell, E. L., M. F. Cohen und J. C. Meeks. 1997. A polyketide-synthase-like gene is involved in the synthesis of heterocyst glycolipids in *Nostoc punctiforme* strain ATCC 29133. *Arch Microbiol* **167**:251-258.
- Castenholz, R. W. 1988. Culturing methods for cyanobacteria. *Cyanobacteria*. Academic Press, San Diego.
- Cavalier-Smith, T. 1998. A revised six-kingdom system of life. *Biol Rev Camb Philos Soc* **73**:203-266.
- Cavalier-Smith, T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int J Syst Evol Microbiol* **52**:297-354.
- Cedergren, R., M. W. Gray, Y. Abel und D. Sankoff. 1988. The evolutionary relationships among known life forms. *J Mol Evol* **28**:98-112.
- Centurion-Lara, A., R. E. LaFond, K. Hevner, C. Godornes, B. J. Molini, W. C. Van Voorhis und S. A. Lukehart. 2004. Gene conversion: a mechanism for generation of heterogeneity in the *tpkK* gene of *Treponema pallidum* during infection. *Mol Microbiol* **52**:1579-1596.
- Challis, G. L., J. Ravel und C. A. Townsend. 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem Biol* **7**:211-224.
- Chang, Z., P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis und D. H. Sherman. 2002. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* **296**:235-247.
- Chang, Z., N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman und W. H. Gerwick. 2004. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* **67**:1356-1367.
- Cheng, Y. Q., G. L. Tang und B. Shen. 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc Natl Acad Sci U S A* **100**:3149-3154.
- Chirala, S. S. und S. J. Wakil. 2004. Structure and function of animal fatty acid synthase. *Lipids* **39**:1045-1053.
- Christiansen, G., E. Dittmann, L. Via Ordorika, R. Rippka, M. Herdman und T. Börner. 2001. Nonribosomal peptide synthetase genes occur in most cyanobacterial genera as evidenced by their distribution in axenic strains of the PCC. *Arch Microbiol* **176**:452-458.
- Cohen, M. F., J. G. Wallis, E. L. Campbell und J. C. Meeks. 1994. Transposon mutagenesis of *Nostoc* sp. strain ATCC 29133, a filamentous cyanobacterium with multiple cellular differentiation alternatives. *Microbiology* **140** (Pt 12):3233-3240.
- Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, 3rd, F. Tekai, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead und B. G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**:537-544.
- Collie, J. N. 1907. Derivatives of the multiple ketene group. *J Chem Soc* **91**:1806-1813.
- Conrad, M. 1990. The geometry of evolution. *BioSystems* **24**.
- Conti, E., T. Stachelhaus, M. A. Marahiel und P. Brick. 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *Embo J* **16**:4174-4183.
- Copley, S. D. 2003. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr Opin Chem Biol* **7**:265-272.
- Cortes, J., S. F. Haydock, G. A. Roberts, D. J. Bevitt und P. F. Leadlay. 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**:176-178.
- Croteau, R., F. Karp, K. C. Wagschal, D. M. Satterwhite, D. C. Hyatt und C. B. Skotland. 1991. Biochemical Characterization of a Spearmint Mutant That Resembles Peppermint in Monoterpene Content. *Plant Physiol* **96**:744-752.

- Dacks, J. B. und W. F. Doolittle. 2001. Reconstructing/deconstructing the earliest eukaryotes: how comparative genomics can help. *Cell* **107**:419-425.
- Damerval, T., G. Guglielmi, J. Houmard und N. T. De Marsac. 1991. Hormogonium Differentiation in the Cyanobacterium *Calothrix*: A Photoregulated Developmental Process. *Plant Cell* **3**:191-201.
- Daubin, V., M. Gouy und G. Perriere. 2001. Bacterial molecular phylogeny using supertree approach. *Genome Inform Ser Workshop Genome Inform* **12**:155-164.
- Davies, J. 1990. What are antibiotics? Archaic functions for modern activities. *Mol. Microbiol.* **4**:1227-1232.
- Davies, J. 2006. Are antibiotics naturally antibiotics? *J Ind Microbiol Biotechnol* **33**:496-499.
- Dembitsky, V. M. und T. Rezanka. 2005. Metabolites produced by nitrogen-fixing *Nostoc* species. *Folia Microbiol (Praha)* **50**:363-391.
- Dittmann, E., B. A. Neilan und T. Börner. 2001. Molecular biology of peptide and polyketide biosynthesis in cyanobacteria. *Appl Microbiol Biotechnol* **57**:467-473.
- Dittmann, E., B. A. Neilan, M. Erhard, H. von Döhren und T. Börner. 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. *Mol Microbiol* **26**:779-787.
- Dittmann, E. und C. Wiegand. 2006. Cyanobacterial toxins--occurrence, biosynthesis and impact on human affairs. *Mol Nutr Food Res* **50**:7-17.
- Dobrindt, U., B. Hochhut, U. Hentschel und J. Hacker. 2004. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* **2**:414-424.
- Donadio, S. und L. Katz. 1992. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in *Saccharopolyspora erythraea*. *Gene* **111**:51-60.
- Donadio, S., M. J. Staver, J. B. McAlpine, S. J. Swanson und L. Katz. 1991. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**:675-679.
- Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. *Science* **284**:2124-2129.
- Drummond, A. und K. Strimmer. 2001. PAL: an object-oriented programming library for molecular evolution and phylogenetics. *Bioinformatics* **17**:662-663.
- Duchaud, E., C. Rusniok, L. Frangeul, C. Buchrieser, A. Givaudan, S. Taourit, S. Bocs, C. Boursaux-Eude, M. Chandler, J. F. Charles, E. Dassa, R. Derosé, S. Derzelle, G. Freyssinet, S. Gaudriault, C. Medigue, A. Lanois, K. Powell, P. Siguier, R. Vincent, V. Wingate, M. Zouine, P. Glaser, N. Boemare, A. Danchin und F. Kunst. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat Biotechnol* **21**:1307-1313.
- Durbin, M. L., G. H. Learn, Jr., G. A. Huttley und M. T. Clegg. 1995. Evolution of the chalcone synthase gene family in the genus *Ipomoea*. *Proc Natl Acad Sci U S A* **92**:3338-3342.
- Eckert, W. A. und J. Kartenbeck. 1997. Proteine: Standardmethoden der Molekular- und Zellbiologie. Springer, Berlin.
- Egan, S., P. Wiener, D. Kallifidas und E. M. Wellington. 1998. Transfer of streptomycin biosynthesis gene clusters within streptomycetes isolated from soil. *Appl Environ Microbiol* **64**:5061-5063.
- Egan, S., P. Wiener, D. Kallifidas und E. M. Wellington. 2001. Phylogeny of *Streptomyces* species and evidence for horizontal transfer of entire and partial antibiotic gene clusters. *Antonie Van Leeuwenhoek* **79**:127-133.
- Eichinger, L., J. A. Pachebat, G. Glockner, M. A. Rajandream, R. Sucgang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu, B. Tunggal, S. Kummerfeld, M. Madera, B. A. Konfortov, F. Rivero, A. T. Bankier, R. Lehmann, N. Hamlin, R. Davies, P. Gaudet, P. Fey, K. Pilcher, G. Chen, D. Saunders, E. Sodergren, P. Davis, A. Kerhornou, X. Nie, N. Hall, C. Anjard, L. Hemphill, N. Bason, P. Farbrother, B. Desany, E. Just, T. Morio, R. Rost, C. Churcher, J. Cooper, S. Haydock, N. van Driessche, A. Cronin, I. Goodhead, D. Muzny, T. Mourier, A. Pain, M. Lu, D. Harper, R. Lindsay, H. Hauser, K. James, M. Quiles, M. Madan Babu, T. Saito, C. Buchrieser, A. Wardroper, M. Felder, M. Thangavelu, D. Johnson, A. Knights, H. Loulseged, K. Mungall, K. Oliver, C. Price, M. A. Quail, H. Urushihara, J. Hernandez, E. Rabinowitsch, D. Steffen, M. Sanders, J. Ma, Y. Kohara, S. Sharp, M. Simmonds, S. Spiegler, A. Tivey, S. Sugano, B. White, D. Walker, J. Woodward, T. Winckler, Y. Tanaka, G. Shaulsky, M. Schleicher, G. Weinstock, A. Rosenthal, E. C. Cox, R. L. Chisholm, R. Gibbs, W. F. Loomis, M. Platzter, R. R. Kay, J. Williams, P. H. Dear, A. A. Noegel, B. Barrell und A. Kuspa. 2005. The genome of the social amoeba *Dictyostelium discoideum*. *Nature* **435**:43-57.
- El-Sayed, A. K., J. Hotherall, S. M. Cooper, E. Stephens, T. J. Simpson und C. M. Thomas. 2003. Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem Biol* **10**:419-430.

- Elhai, J. und C. P. Wolk. 1988. Conjugal transfer of DNA to cyanobacteria. *Methods Enzymol* **167**:747-754.
- Embley, T. M. und W. Martin. 2006. Eukaryotic evolution, changes and challenges. *Nature* **440**:623-630.
- Eppelmann, K., T. Stachelhaus und M. A. Marahiel. 2002. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry* **41**:9718-9726.
- Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, MA (USA).
- Felsenstein, J. 2002. PHYLIP (phylogeny inference package). Version 3.6. Distributed by the author.
- Fernandes, N. D. und P. E. Kolattukudy. 1996. Cloning, sequencing and characterization of a fatty acid synthase-encoding gene from *Mycobacterium tuberculosis* var. *bovis* BCG. *Gene* **170**:95-99.
- Ferrer, J. L., J. M. Jez, M. E. Bowman, R. A. Dixon und J. P. Noel. 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Biol* **6**:775-784.
- Figurski, D. H. und D. R. Helinski. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc Natl Acad Sci U S A* **76**:1648-1652.
- Firm, R. D. und C. G. Jones. 2000. The evolution of secondary metabolism - a unifying model. *Mol Microbiol* **37**:989-994.
- Firm, R. D. und C. G. Jones. 2003. Natural products--a simple model to explain chemical diversity. *Nat Prod Rep* **20**:382-391.
- Firm, R. D. und C. G. Jones. 2006a. Do we need a new hypothesis to explain plant VOC emissions? *Trends Plant Sci* **11**:112-113; author reply 113-114.
- Firm, R. D. und C. G. Jones. 2006b. Response to Pichersky et al.: Correcting a misconception about the screening hypothesis. *Trends Plant Sci* **11**:422.
- Forkmann, G. und W. Heller. 1999. Biosynthesis of flavonoids. Pp. 713-748 in U. Sankawa, ed. *Comprehensive natural products chemistry*. Elsevier, Oxford.
- Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen und C. R. Woese. 1980. The phylogeny of prokaryotes. *Science* **209**:457-463.
- Fox, J. L. und F. Lynen. 1980. Characterization of the flavoenzyme enoyl reductase of fatty acid synthetase from yeast. *Eur J Biochem* **109**:417-424.
- Fraenkel, G. S. 1959. The raison d'etre of secondary plant substances. *Science* **129**:1466-1470.
- Frey, M., C. Stettner, P. W. Pare, E. A. Schmelz, J. H. Tumlinson und A. Gierl. 2000. An herbivore elicitor activates the gene for indole emission in maize. *Proc Natl Acad Sci U S A* **97**:14801-14806.
- Funa, N., Y. Ohnishi, I. Fujii, M. Shibuya, Y. Ebizuka und S. Horinouchi. 1999. A new pathway for polyketide synthesis in microorganisms. *Nature* **400**:897-899.
- Gaisser, S., A. Trefzer, S. Stockert, A. Kirschning und A. Bechthold. 1997. Cloning of an avilamycin biosynthetic gene cluster from *Streptomyces viridochromogenes* Tu57. *J Bacteriol* **179**:6271-6278.
- Gaitatzis, N., A. Hans, R. Müller und S. Beyer. 2001. The *mtaA* gene of the myxothiazol biosynthetic gene cluster from *Stigmatella aurantiaca* DW4/3-1 encodes a phosphopantetheinyl transferase that activates polyketide synthases and polypeptide synthetases. *J Biochem (Tokyo)* **129**:119-124.
- Gaitatzis, N., B. Silakowski, B. Kunze, G. Nordsiek, H. Blocker, G. Hofle und R. Müller. 2002. The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J Biol Chem* **277**:13082-13090.
- Gamielidien, J., A. Ptitsyn und W. Hide. 2002. Eukaryotic genes in *Mycobacterium tuberculosis* could have a role in pathogenesis and immunomodulation. *Trends Genet* **18**:5-8.
- Geck, M. K. und J. F. Kirsch. 1999. A novel, definitive test for substrate channeling illustrated with the aspartate aminotransferase/malate dehydrogenase system. *Biochemistry* **38**:8032-8037.
- Gehring, A. M., I. I. Mori, R. D. Perry und C. T. Walsh. 1998. The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of *Yersinia pestis*. *Biochemistry* **37**:17104.
- Gehring, A. M., S. T. Wang, D. B. Kearns, N. Y. Storer und R. Losick. 2004. Novel genes that influence development in *Streptomyces coelicolor*. *J Bacteriol* **186**:3570-3577.
- Gevers, W., H. Kleinkauf und F. Lipmann. 1968. The activation of amino acids for biosynthesis of gramicidin S. *Proc Natl Acad Sci U S A* **60**:269-276.
- Gevers, W., H. Kleinkauf und F. Lipmann. 1969. Peptidyl transfers in gramicidin S biosynthesis from enzyme-bound thioester intermediates. *Proc Natl Acad Sci U S A* **63**:1335-1342.

- Ginolhac, A., C. Jarrin, P. Robe, G. Perriere, T. M. Vogel, P. Simonet und R. Nalin. 2005. Type I polyketide synthases may have evolved through horizontal gene transfer. *J Mol Evol* **60**:716-725.
- Givan, C. V. 1983. The source of acetyl coenzyme A in chloroplasts of higher plants. *Physiol Plant* **57**:311-316.
- Gocht, M. und M. A. Marahiel. 1994. Analysis of core sequences in the D-Phe activating domain of the multifunctional peptide synthetase TycA by site-directed mutagenesis. *J Bacteriol* **176**:2654-2662.
- Goh, E. B., G. Yim, W. Tsui, J. McClure, M. G. Surette und J. Davies. 2002. Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc Natl Acad Sci U S A* **99**:17025-17030.
- Gokhale, R. S., S. Y. Tsuji, D. E. Cane und C. Khosla. 1999. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**:482-485.
- Golakoti, T., J. Ogino, C. E. Heltzel, T. Le Husebo, C. M. Jensen, L. K. Larsen, G. M. L. Patterson, R. E. Moore, S. L. Mooberry, T. H. Corbett und F. A. Valeriores. 1995. Structure determination, conformational analysis, chemical stability studies und antitumor evaluation of the cryptophycins. Isolation of 18 new analogs from *Nostoc* sp. strain GSV 224t. *J Am Chem Soc* **117**:12030-12049.
- Golakoti, T., W. Y. Yoshida, S. Chaganty und R. E. Moore. 2000. Isolation and structures of nostopeptolides A1, A2 und A3 from the cyanobacterium *Nostoc* sp. GSV224. *Tetrahedron* **56**:9093-9102.
- Golakoti, T., W. Y. Yoshida, S. Chaganty und R. E. Moore. 2001. Isolation and structure determination of nostocyclopeptides A1 und A2 from the terrestrial cyanobacterium *Nostoc* sp. ATCC53789. *J Nat Prod* **64**:54-59.
- Goldman, B. S., W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. A. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand und H. B. Kaplan. 2006. Evolution of sensory complexity recorded in a myxobacterial genome. *Proc Natl Acad Sci U S A* **103**:15200-15205.
- Gonnet, G. H., M. A. Cohen und S. A. Benner. 1992. Exhaustive matching of the entire protein sequence database. *Science* **256**:1443-1445.
- Gonzalez-Lergier, J., L. J. Broadbelt und V. Hatzimanikatis. 2005. Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways. *J Am Chem Soc* **127**:9930-9938.
- Gottschalk, G. 1986. *Bacterial Metabolism*. Spriner, New York.
- Graur, D. und W.-H. Li. 1999. *Fundamentals of molecular evolution*. Sinauer, Sunderland, MA (USA).
- Gray, M. W., G. Burger und B. F. Lang. 1999. Mitochondrial evolution. *Science* **283**:1476-1481.
- Griffiths, A. D. und D. S. Tawfik. 2000. Man-made enzymes--from design to in vitro compartmentalisation. *Curr Opin Biotechnol* **11**:338-353.
- Gross, G. G. 1981. *The biochemistry of plants*. Academic Press, London.
- Guenzi, E., G. Galli, I. Grgurina, D. C. Gross und G. Grandi. 1998. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic und eukaryotic peptide synthetases. *J Biol Chem* **273**:32857-32863.
- Guindon, S. und O. Gascuel. 2003. A simple, fast und accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**:696-704.
- Han, L., S. Lobo und K. A. Reynolds. 1998. Characterization of beta-ketoacyl-acyl carrier protein synthase III from *Streptomyces glaucescens* und its role in initiation of fatty acid biosynthesis. *J Bacteriol* **180**:4481-4486.
- Hardt, I. H., H. Steinmetz, K. Gerth, F. Sasse, H. Reichenbach und G. Hofle. 2001. New natural epothilones from *Sorangium cellulosum*, strains So ce90/B2 und So ce90/D13: isolation, structure elucidation und SAR studies. *J Nat Prod* **64**:847-856.
- Hartmann, T. 2004. Plant-derived secondary metabolites as defensive chemicals in herbivorous insects: a case study in chemical ecology. *Planta* **219**:1-4.
- Haslam, E. 1989. *Plant polyphenols. Vegetable tannins revisited*. Cambridge University Press, Cambridge.
- He, J. und C. Hertweck. 2003. Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem Biol* **10**:1225-1232.
- Heath, R. J. und C. O. Rock. 1996. Inhibition of beta-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in *Escherichia coli*. *J Biol Chem* **271**:10996-11000.

- Helariutta, Y., M. Kotilainen, P. Elomaa, N. Kalkkinen, K. Bremer, T. H. Teeri und V. A. Albert. 1996. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: evolution with substrate change and catalytic simplification. *Proc Natl Acad Sci U S A* **93**:9033-9038.
- Heller, A. und T. Koch. 1998. [Pharmacologic aspects of polyunsaturated fatty acids in parenteral nutrition]. *Anesthesiol Intensivmed Notfallmed Schmerzther* **33**:77-87.
- Henikoff, S. und J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* **89**:10915-10919.
- Henriquez, F. L., T. A. Richards, F. Roberts, R. McLeod und C. W. Roberts. 2005. The unusual mitochondrial compartment of *Cryptosporidium parvum*. *Trends Parasitol* **21**:68-74.
- Herdman, M. und R. Rippka. 1988. Cellular differentiation: hormogonia and baeocytes. Pp. 232-242. *Methods in Enzymology*.
- Hilker, M., C. Kobs, M. Varama und K. Schrank. 2002. Insect egg deposition induces *Pinus sylvestris* to attract egg parasitoids. *J Exp Biol* **205**:455-461.
- Hinrichs, W., C. Kisker, M. Duvel, A. Müller, K. Tovar, W. Hillen und W. Saenger. 1994. Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science* **264**:418-420.
- Hitchman, T. S., J. Crosby, K. J. Byrom, R. J. Cox und T. J. Simpson. 1998. Catalytic self-acylation of type II polyketide synthase acyl carrier proteins. *Chem Biol* **5**:35-47.
- Hoffmann, D., J. M. Hevel, R. E. Moore und B. S. Moore. 2003. Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. *Gene* **311**:171-180.
- Holak, T. A., M. Nilges, J. H. Prestegard, A. M. Gronenborn und G. M. Clore. 1988. Three-dimensional structure of acyl carrier protein in solution determined by nuclear magnetic resonance and the combined use of dynamical simulated annealing and distance geometry. *Eur J Biochem* **175**:9-15.
- Holder, M. und P. O. Lewis. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat Rev Genet* **4**:275-284.
- Hopwood, D. A. 1997. Genetic Contributions to Understanding Polyketide Synthases. *Chem Rev* **97**:2465-2498.
- Hopwood, D. A. und D. H. Sherman. 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Annu Rev Genet* **24**:37-66.
- Hranueli, D., J. Cullum, B. Basrak, P. Goldstein und P. F. Long. 2005. Plasticity of the streptomyces genome-evolution and engineering of new antibiotics. *Curr Med Chem* **12**:1697-1704.
- Huang, J., N. Mullapudi, C. A. Lancto, M. Scott, M. S. Abrahamsen und J. C. Kissinger. 2004. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol* **5**:R88.
- Huang, W., J. Jia, P. Edwards, K. Dehesh, G. Schneider und Y. Lindqvist. 1998. Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from *E.coli* reveals the molecular architecture of condensing enzymes. *Embo J* **17**:1183-1191.
- Hughes, A. L. 1994. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* **256**:119-124.
- Hunsucker, S. W., K. Klage, S. M. Slaughter, M. Potts und R. F. Helm. 2004. A preliminary investigation of the *Nostoc punctiforme* proteome. *Biochem Biophys Res Commun* **317**:1121-1127.
- Hutchinson, C. R. und I. Fujii. 1995. Polyketide synthase gene manipulation: a structure-function approach in engineering novel antibiotics. *Annu Rev Microbiol* **49**:201-238.
- Ikeda, H., J. Ishikawa, A. Hanamoto, M. Shinose, H. Kikuchi, T. Shiba, Y. Sakaki, M. Hattori und S. Omura. 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**:526-531.
- Ikeda, H. und S. Omura. 1997. Avermectin Biosynthesis. *Chem Rev* **97**:2591-2610.
- Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa und T. Miyata. 1989. Evolutionary relationship of archaeobacteria, eubacteria und eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci U S A* **86**:9355-9359.
- Izumikawa, M., P. R. Shipley, J. N. Hopke, T. O'Hare, L. Xiang, J. P. Noel und B. S. Moore. 2003. Expression and characterization of the type III polyketide synthase 1,3,6,8-tetrahydroxynaphthalene synthase from *Streptomyces coelicolor* A3(2). *J Ind Microbiol Biotechnol* **30**:510-515.

- Jackowski, S. und C. O. Rock. 1987. Acetoacetyl-acyl carrier protein synthase, a potential regulator of fatty acid biosynthesis in bacteria. *J Biol Chem* **262**:7927-7931.
- Jakobi, C., L. Oberer, C. Quiquerez, W. A. König und J. Weckesser. 1995. Cyanopeptolin S, a sulfate-containing depsipeptide from a water bloom of *Microcystis* sp. *FEMS Microbiol Lett* **129**:129-133.
- Jenni, S., M. Leibundgut, T. Maier und N. Ban. 2006. Architecture of a fungal fatty acid synthase at 5 Å resolution. *Science* **311**:1263-1267.
- Jensen, R. A. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**:409-425.
- Jez, J. M., J. L. Ferrer, M. E. Bowman, M. B. Austin, J. Schroder, R. A. Dixon und J. P. Noel. 2001. Structure and mechanism of chalcone synthase-like polyketide synthases. *J Ind Microbiol Biotechnol* **27**:393-398.
- Jones, C. G. und R. D. Finn. 1991. On the evolution of plant secondary metabolite chemical diversity. *Phil. Trans. R. Soc. Lon. B.* **333**:273-280.
- Jones, D. T., W. R. Taylor und J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275-282.
- Jordan, I. K., K. S. Makarova, Y. I. Wolf und E. V. Koonin. 2001. Gene conversions in genes encoding outer-membrane proteins in *H. pylori* and *C. pneumoniae*. *Trends Genet* **17**:7-10.
- Joshi, A. K., V. S. Rangan, A. Witkowski und S. Smith. 2003. Engineering of an active animal fatty acid synthase dimer with only one competent subunit. *Chem Biol* **10**:169-173.
- Kao, C. M., R. Pieper, D. E. Cane und C. Khosla. 1996. Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. *Biochemistry* **35**:12363-12368.
- Kasahara, T. und T. Kato. 2003. Nutritional biochemistry: A new redox-cofactor vitamin for mammals. *Nature* **422**:832.
- Kaulmann, U. und C. Hertweck. 2002. Biosynthesis of polyunsaturated fatty acids by polyketide synthases. *Angew Chem Int Ed Engl* **41**:1866-1869.
- Keeling, P. J., G. Burger, D. G. Durnford, B. F. Lang, R. W. Lee, R. E. Pearlman, A. J. Roger und M. W. Gray. 2005. The tree of eukaryotes. *Trends Ecol Evol* **20**:670-676.
- Kehr, J. C., Y. Zilliges, A. Springer, M. D. Disney, D. D. Ratner, C. Bouchier, P. H. Seeberger, N. T. de Marsac und E. Dittmann. 2006. A mannan binding lectin is involved in cell-cell attachment in a toxic strain of *Microcystis aeruginosa*. *Mol Microbiol* **59**:893-906.
- Keller, N. P., G. Turner und J. W. Bennett. 2005. Fungal secondary metabolism - from biochemistry to genomics. *Nat Rev Microbiol* **3**:937-947.
- Kepler, T. B. 1997. Codon bias and plasticity in immunoglobulins. *Mol Biol Evol* **14**:637-643.
- Kessler, A. und I. T. Baldwin. 2002. Plant responses to insect herbivory: the emerging molecular analysis. *Annu Rev Plant Biol* **53**:299-328.
- Khersonsky, O., C. Roodveldt und D. S. Tawfik. 2006. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* **10**:498-508.
- Khosla, C. und P. B. Harbury. 2001. Modular enzymes. *Nature* **409**:247-252.
- Kikuchi, S., D. L. Rainwater und P. E. Kolattukudy. 1992. Purification and characterization of an unusually large fatty acid synthase from *Mycobacterium tuberculosis* var. *bovis* BCG. *Arch Biochem Biophys* **295**:318-326.
- Kinsella, R. J., D. A. Fitzpatrick, C. J. Creevey und J. O. McNerney. 2003. Fatty acid biosynthesis in *Mycobacterium tuberculosis*: lateral gene transfer, adaptive evolution und gene duplication. *Proc Natl Acad Sci U S A* **100**:10320-10325.
- Kinsella, R. J. und J. O. McNerney. 2003. Eukaryotic genes in *Mycobacterium tuberculosis*? Possible alternative explanations. *Trends Genet* **19**:687-689.
- Kirschner, M. und J. Gerhart. 1998. Evolvability. *Proc Natl Acad Sci U S A* **95**:8420-8427.
- Kishino, H. und M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data und the branching order in hominoidea. *J Mol Evol* **29**:170-179.
- Kline, K. A., E. V. Sechman, E. P. Skaar und H. S. Seifert. 2003. Recombination, repair und replication in the pathogenic *Neisseriae*: the 3 R's of molecular genetics of two human-specific bacterial pathogens. *Mol Microbiol* **50**:3-13.
- Knoll, A. H. 2003. The geological consequences of evolution. *Geobiology* **1**:3-14.
- Kodumal, S. J., K. G. Patel, R. Reid, H. G. Menzella, M. Welch und D. V. Santi. 2004. Total synthesis of long DNA sequences: synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proc Natl Acad Sci U S A* **101**:15573-15578.

- Konz, D., A. Klens, K. Schorgendorfer und M. A. Marahiel. 1997. The bacitracin biosynthesis operon of *Bacillus licheniformis* ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* **4**:927-937.
- Koonin, E. V., K. S. Makarova und L. Aravind. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**:709-742.
- Kossel, A. 1891. Über die chemische Zusammensetzung der Zelle. *Arch Physiol*:181.
- Kumar, S., K. Tamura, I. B. Jakobsen und M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**:1244-1245.
- Kumar, S., K. Tamura und M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**:150-163.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685.
- Lang, B. F., M. W. Gray und G. Burger. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* **33**:351-397.
- Lau, J., H. Fu, D. E. Cane und C. Khosla. 1999. Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry* **38**:1643-1651.
- Lauritzen, L., H. S. Hansen, M. H. Jorgensen und K. F. Michaelsen. 2001. The essentiality of long chain n-3 fatty acids in relation to development and function of the brain and retina. *Prog Lipid Res* **40**:1-94.
- Lautru, S. und G. L. Challis. 2004. Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology* **150**:1629-1636.
- Lautru, S., R. J. Deeth, L. M. Bailey und G. L. Challis. 2005. Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat Chem Biol* **1**:265-269.
- Leesong, M., B. S. Henderson, J. R. Gillig, J. M. Schwab und J. L. Smith. 1996. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**:253-264.
- Leigh Brown, A. J. und D. Ish-Horowicz. 1981. Evolution of the 87A and 87C heat-shock loci in *Drosophila*. *Nature* **290**:677-682.
- Liao, D. 2000. Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J Mol Evol* **51**:305-317.
- Liebhauer, S. A., M. Goossens und Y. W. Kan. 1981. Homology and concerted evolution at the alpha 1 and alpha 2 loci of human alpha-globin. *Nature* **290**:26-29.
- Liebig, J. 1843. *Handbuch der Chemie mit Rücksicht auf Pharmacie*, Heidelberg.
- Linne, U. und M. A. Marahiel. 2000. Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry* **39**:10439-10447.
- Linne, U., D. B. Stein, H. D. Mootz und M. A. Marahiel. 2003. Systematic and quantitative analysis of protein-protein recognition between nonribosomal peptide synthetases investigated in the tyrocidine biosynthetic template. *Biochemistry* **42**:5114-5124.
- Lio, P. und N. Goldman. 1998. Models of molecular evolution and phylogeny. *Genome Res* **8**:1233-1244.
- Liu, H., L. Campbell und H. A. Nolla. 1997. *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat Microb Ecol* **12**:39-47.
- Liu, W., S. D. Christenson, S. Standage und B. Shen. 2002. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297**:1170-1173.
- Liu, W., K. Nonaka, L. Nie, J. Zhang, S. D. Christenson, J. Bae, S. G. Van Lanen, E. Zazopoulos, C. M. Farnet, C. F. Yang und B. Shen. 2005. The neocarzinostatin biosynthetic gene cluster from *Streptomyces carzinostaticus* ATCC 15944 involving two iterative type I polyketide synthases. *Chem Biol* **12**:293-302.
- Luckner, M. 1990. *Secondary Metabolites in Microorganisms, Plants und Animals*. Gustav Fischer Verlag, Jena.
- Luesch, H., D. Hoffmann, J. M. Hevel, J. E. Becker, T. Golakoti und R. E. Moore. 2003. Biosynthesis of 4-methylproline in cyanobacteria: cloning of nosE and nosF genes and biochemical characterization of the encoded dehydrogenase and reductase activities. *J Org Chem* **68**:83-91.
- Lynch, M. und J. S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151-1155.

- Lynen, F. 1969. Yeast fatty acid synthase. *Methods Enzymol* **14**:17-32.
- Maddison, W. R. und W. P. Maddison. 2000. MacClade version 4.0 in S. Associates, ed, Sunderland, MA.
- Magarvey, N. A., Z. Q. Beck, T. Golakoti, Y. Ding, U. Huber, T. K. Hemscheidt, D. Abelson, R. E. Moore und D. H. Sherman. 2006. Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts. *ACS Chem Biol* **1**:766-779.
- Maier, T., S. Jenni und N. Ban. 2006. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311**:1258-1262.
- Malik, V. S. 1980. Microbial secondary metabolism. *Trends Biochem Sci* **5**:68-72.
- Marahiel, M. A., T. Stachelhaus und H. D. Mootz. 1997. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev* **97**:2651-2674.
- Marrakchi, H., Y. M. Zhang und C. O. Rock. 2002. Mechanistic diversity and regulation of Type II fatty acid synthesis. *Biochem Soc Trans* **30**:1050-1055.
- Martin, W., C. Rotte, M. Hoffmeister, U. Theissen, G. Gelius-Dietrich, S. Ahr und K. Henze. 2003. Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?) und a tree of genomes revisited. *IUBMB Life* **55**:193-204.
- Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa und D. Penny. 2002. Evolutionary analysis of Arabidopsis, cyanobacterial und chloroplast genomes reveals plastid phylogeny und thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A* **99**:12246-12251.
- Martin, W., B. Stoebe, V. Goremykin, S. Hapsmann, M. Hasegawa und K. V. Kowallik. 1998. Gene transfer to the nucleus und the evolution of chloroplasts. *Nature* **393**:162-165.
- Matharu, A. L., R. J. Cox, J. Crosby, K. J. Byrom und T. J. Simpson. 1998. MCAT is not required for in vitro polyketide synthesis in a minimal actinorhodin polyketide synthase from *Streptomyces coelicolor*. *Chem Biol* **5**:699-711.
- McDaniel, R., S. Ebert-Khosla, D. A. Hopwood und C. Khosla. 1993. Engineered biosynthesis of novel polyketides. *Science* **262**:1546-1550.
- McDaniel, R., A. Thamchaipenet, C. Gustafsson, H. Fu, M. Betlach und G. Ashley. 1999. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc Natl Acad Sci U S A* **96**:1846-1851.
- McFadden, G. I. 1999. Plastids und protein targeting. *J Eukaryot Microbiol* **46**:339-346.
- McGuffin, L. J., K. Bryson und D. T. Jones. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**:404-405.
- Meeks, J. C. 1998. Symbiosis between nitrogen-fixing cyanobacteria und plants. *BioScience* **48**:266-276.
- Meeks, J. C., E. L. Campbell, M. L. Summers und F. C. Wong. 2002. Cellular differentiation in the cyanobacterium *Nostoc punctiforme*. *Arch Microbiol* **178**:395-403.
- Meeks, J. C. und J. Elhai. 2002. Regulation of cellular differentiation in filamentous cyanobacteria in free-living und plant-associated symbiotic growth states. *Microbiol Mol Biol Rev* **66**:94-121; table of contents.
- Meeks, J. C., J. Elhai, T. Thiel, M. Potts, F. Larimer, J. Lamerdin, P. Predki und R. Atlas. 2001. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. *Photosynth Res* **70**:85-106.
- Meiser, P., H. B. Bode und R. Müller. 2006. The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc Natl Acad Sci U S A* **103**:19128-19133.
- Menzella, H. G., R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. A. Hopwood und D. V. Santi. 2005. Combinatorial polyketide biosynthesis by de novo design und rearrangement of modular polyketide synthase genes. *Nat Biotechnol* **23**:1171-1176.
- Metsä-Ketelä, M., L. Halo, E. Munukka, J. Hakala, P. Mantsala und K. Ylihonko. 2002. Molecular evolution of aromatic polyketides und comparative sequence analysis of polyketide ketosynthase und 16S ribosomal DNA genes from various streptomyces species. *Appl Environ Microbiol* **68**:4472-4479.
- Metsä-Ketelä, M., K. Ylihonko und P. Mantsala. 2004. Partial activation of a silent angucycline-type gene cluster from a rubromycin beta producing *Streptomyces* sp. PGA64. *J Antibiot (Tokyo)* **57**:502-510.
- Metz, J. G., P. Roessler, D. Facciotti, C. Levering, F. Dittrich, M. Lassner, R. Valentine, K. Lardizabal, F. Domergue, A. Yamada, K. Yazawa, V. Knauf und J. Browse. 2001. Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes und eukaryotes. *Science* **293**:290-293.

- Meurer, G., G. Biermann, A. Schutz, S. Harth und E. Schweizer. 1992. Molecular structure of the multifunctional fatty acid synthetase gene of *Brevibacterium ammoniagenes*: its sequence of catalytic domains is formally consistent with a head-to-tail fusion of the two yeast genes FAS1 and FAS2. *Mol Gen Genet* **232**:106-116.
- Mikalsen, B., G. Boison, O. M. Skulberg, J. Fastner, W. Davies, T. M. Gabrielsen, K. Rudi und K. S. Jakobsen. 2003. Natural variation in the microcystin synthetase operon *mcvABC* and impact on microcystin production in *Microcystis* strains. *J Bacteriol* **185**:2774-2785.
- Miller, D. A., L. Luo, N. Hillson, T. A. Keating und C. T. Walsh. 2002. Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chem Biol* **9**:333-344.
- Miller, D. A., C. T. Walsh und L. Luo. 2001. C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase. *J Am Chem Soc* **123**:8434-8435.
- Minnikin, D. E. 1982. Lipids: Complex lipids, their chemistry, biosynthesis and roles. Pp. 95-184 in C. Ratledge und J. L. Stanford, eds. *The biology of the mycobacteria*. Academic, London.
- Mira, A., H. Ochman und N. A. Moran. 2001. Deletional bias and the evolution of bacterial genomes. *Trends Genet* **17**:589-596.
- Misawa, N., Y. Satomi, K. Kondo, A. Yokoyama, S. Kajiwar, T. Saito, T. Ohtani und W. Miki. 1995. Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J Bacteriol* **177**:6575-6584.
- Moche, M., G. Schneider, P. Edwards, K. Dehesh und Y. Lindqvist. 1999. Structure of the complex between the antibiotic cerulenin and its target, beta-ketoacyl-acyl carrier protein synthase. *J Biol Chem* **274**:6031-6034.
- Mochizuki, S., K. Hiratsu, M. Suwa, T. Ishii, F. Sugino, K. Yamada und H. Kinashi. 2003. The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol* **48**:1501-1510.
- Moffitt, M. C. und B. A. Neilan. 2003. Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *J Mol Evol* **56**:446-457.
- Montoya, J. P., C. M. Holl, J. P. Zehr, A. Hansen, T. A. Villareal und D. G. Capone. 2004. High rates of N₂ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430**:1027-1032.
- Moore, B. S. und C. Hertweck. 2002. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat Prod Rep* **19**:70-99.
- Moore, B. S. und J. N. Hopke. 2001. Discovery of a new bacterial polyketide biosynthetic pathway. *ChemBiochem* **2**:35-38.
- Moore, R. E., J. L. Chen, B. S. Moore, G. M. L. Patterson und W. W. Carmichael. 1991. Biosynthesis of microcystin-LR. Origin of the carbons in the Adda and Masp units. *J Am Chem Soc* **113**:5083-5084.
- Mootz, H. D., R. Finking und M. A. Marahiel. 2001. 4'-phosphopantetheine transfer in primary and secondary metabolism of *Bacillus subtilis*. *J Biol Chem* **276**:37289-37298.
- Moss, S. J., C. J. Martin und B. Wilkinson. 2004. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat Prod Rep* **21**:575-593.
- Müller, R. 2004. Don't classify polyketide synthases. *Chem Biol* **11**:4-6.
- Nagyilaki, T. 1984. Evolution of multigene families under interchromosomal gene conversion. *Proc Natl Acad Sci U S A* **81**:3796-3800.
- Nagyilaki, T. und T. D. Petes. 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* **100**:315-337.
- Nei, M. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Nei, M. und A. P. Rooney. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39**:121-152.
- Nichols, D., J. Bowman, K. Sanderson, C. M. Nichols, T. Lewis, T. McMeekin und P. D. Nichols. 1999. Developments with antarctic microorganisms: culture collections, bioactivity screening, taxonomy, PUFA production and cold-adapted enzymes. *Curr Opin Biotechnol* **10**:240-246.
- Nielsen, R. und andere. 2005. *Statistical methods in molecular evolution*. Springer, New York.
- O'Brien, P. J. und D. Herschlag. 1999. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* **6**:R91-R105.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer, New York.
- Ohta, T. 1990. How gene families evolve. *Theor Pop Biol* **37**:213-219.

- Ohta, T. 1998. On the pattern of polymorphisms at major histocompatibility complex loci. *J Mol Evol* **46**:633-638.
- Oliynyk, M., M. J. Brown, J. Cortes, J. Staunton und P. F. Leadlay. 1996. A hybrid modular polyketide synthase obtained by domain swapping. *Chem Biol* **3**:833-839.
- Olsen, J. G., A. Kadziola, P. von Wettstein-Knowles, M. Siggaard-Andersen und S. Larsen. 2001. Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **9**:233-243.
- Omura, S., H. Ikeda, J. Ishikawa, A. Hanamoto, C. Takahashi, M. Shinose, Y. Takahashi, H. Horikawa, H. Nakazawa, T. Osonoe, H. Kikuchi, T. Shiba, Y. Sakaki und M. Hattori. 2001. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci U S A* **98**:12215-12220.
- Orjala, J. und W. H. Gerwick. 1996. Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* **59**:427-430.
- Orth, P., D. Schnappinger, W. Hillen, W. Saenger und W. Hinrichs. 2000. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat Struct Biol* **7**:215-219.
- Owen, S. M. und J. Penuelas. 2005. Opportunistic emissions of volatile isoprenoids. *Trends Plant Sci* **10**:420-426.
- Paitan, Y., G. Alon, E. Orr, E. Z. Ron und E. Rosenberg. 1999. The first gene in the biosynthesis of the polyketide antibiotic TA of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase. *J Mol Biol* **286**:465-474.
- Pao, G. M. und M. H. Saier, Jr. 1995. Response regulators of bacterial signal transduction systems: selective domain shuffling during evolution. *J Mol Evol* **40**:136-154.
- Partensky, F., W. R. Hess und D. Vaulot. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**:106-127.
- Patten, P. A., R. J. Howard und W. P. Stemmer. 1997. Applications of DNA shuffling to pharmaceuticals and vaccines. *Curr Opin Biotechnol* **8**:724-733.
- Pavela-Vrancic, M., R. Dieckmann und H. von Döhren. 2004. ATPase activity of non-ribosomal peptide synthetases. *Biochim Biophys Acta* **1696**:83-91.
- Pereto, J., P. Lopez-Garcia und D. Moreira. 2004. Ancestral lipid biosynthesis and early membrane evolution. *Trends Biochem Sci* **29**:469-477.
- Peters, N. K., J. W. Frost und S. R. Long. 1986. A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science* **233**:977-980.
- Pfeifer, B. A. und C. Khosla. 2001. Biosynthesis of polyketides in heterologous hosts. *Microbiol Mol Biol Rev* **65**:106-118.
- Pfeifer, V., G. J. Nicholson, J. Ries, J. Recktenwald, A. B. Schefer, R. M. Shawky, J. Schroder, W. Wohlleben und S. Pelzer. 2001. A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine. *J Biol Chem* **276**:38370-38377.
- Philippe, H. und C. J. Douady. 2003. Horizontal gene transfer and phylogenetics. *Curr Opin Microbiol* **6**:498-505.
- Pichersky, E., T. D. Sharkey und J. Gershenzon. 2006. Plant volatiles: a lack of function or a lack of knowledge? *Trends Plant Sci* **11**:421; author reply 422-423.
- Piel, J. 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc Natl Acad Sci U S A* **99**:14002-14007.
- Piel, J., D. Hui, N. Fusetani und S. Matsunaga. 2004. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ Microbiol* **6**:921-927.
- Plotkin, J. B. und J. Dushoff. 2003. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proc Natl Acad Sci U S A* **100**:7152-7157.
- Price, A. C., Y. M. Zhang, C. O. Rock und S. W. White. 2001. Structure of beta-ketoacyl-[acyl carrier protein] reductase from *Escherichia coli*: negative cooperativity and its structural basis. *Biochemistry* **40**:12772-12781.
- Qiu, X., C. A. Janson, A. K. Konstantinidis, S. Nwagwu, C. Silverman, W. W. Smith, S. Khandekar, J. Lonsdale und S. S. Abdel-Meguid. 1999. Crystal structure of beta-ketoacyl-acyl carrier protein synthase III. A key condensing enzyme in bacterial fatty acid biosynthesis. *J Biol Chem* **274**:36465-36471.

- Ragan, M. A. 2001. Detection of lateral gene transfer among microbial genomes. *Curr Opin Genet Dev* **11**:620-626.
- Rainey, P. B. und T. F. Cooper. 2004. Evolution of bacterial diversity and the origins of modularity. *Res Microbiol* **155**:370-375.
- Ramos, J. L., M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan und R. Tobes. 2005. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* **69**:326-356.
- Rangan, V. S., A. K. Joshi und S. Smith. 2001. Mapping the functional topology of the animal fatty acid synthase by mutant complementation in vitro. *Biochemistry* **40**:10792-10799.
- Rausch, C., T. Weber, O. Kohlbacher, W. Wohlleben und D. H. Huson. 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res* **33**:5799-5808.
- Rawlings, B. J. 1999. Biosynthesis of polyketides (other than actinomycete macrolides). *Nat Prod Rep* **16**:425-484.
- Reeves, C. D. 2003. The enzymology of combinatorial biosynthesis. *Crit Rev Biotechnol* **23**:95-147.
- Reichenbach, H. 1999. The ecology of the myxobacteria. *Environ Microbiol* **1**:15-21.
- Reimann, C., H. M. Patel, L. Serino, M. Barone, C. T. Walsh und D. Haas. 2001. Essential PchG-dependent reduction in pyochelin biosynthesis of *Pseudomonas aeruginosa*. *J Bacteriol* **183**:813-820.
- Reuter, K., M. R. Mofid, M. A. Marahiel und R. Ficner. 1999. Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily. *Embo J* **18**:6823-6831.
- Revill, W. P., M. J. Bibb und D. A. Hopwood. 1995. Purification of a malonyltransferase from *Streptomyces coelicolor* A3(2) and analysis of its genetic determinant. *J Bacteriol* **177**:3946-3952.
- Richardson, M. und C. Khosla. 1999. Structure, function und engineering of bacterial aromatic polyketide synthases. Pp. 473-494 in U. Sankawa, ed. *Polyketides and other secondary metabolites including fatty acids and their derivatives*. Elsevier, Oxford.
- Richardt, A., T. Kemme, S. Wagner, D. Schwarzer, M. A. Marahiel und B. T. Hovemann. 2003. Ebony, a novel nonribosomal peptide synthetase for beta-alanine conjugation with biogenic amines in *Drosophila*. *J Biol Chem* **278**:41160-41166.
- Ridley, M. 2004. *Evolution*. Blackwell Publishing, Malden, MA (USA).
- Rippka, R., J. Dornelles, J. B. Waterbury, M. Herdman und R. Y. Stanier. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**:1-61.
- Rock, C. O. und S. Jackowski. 2002. Forty years of bacterial fatty acid synthesis. *Biochem Biophys Res Commun* **292**:1155-1166.
- Rodriguez, C. und D. Romero. 1998. Multiple recombination events maintain sequence identity among members of the nitrogenase multigene family in *Rhizobium etli*. *Genetics* **149**:785-794.
- Ronquist, F. und J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572-1574.
- Ronquist, F., J. P. Huelsenbeck und P. van der Mark. 2005. MrBayes 3.1 Manual. Distribution durch die Autoren.
- Ruan, X., A. Pereda, D. L. Stassi, D. Zeidner, R. G. Summers, M. Jackson, A. Shivakumar, S. Kakavas, M. J. Staver, S. Donadio und L. Katz. 1997. Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *J Bacteriol* **179**:6416-6425.
- Russell, N. J. und D. S. Nichols. 1999. Polyunsaturated fatty acids in marine bacteria--a dogma rewritten. *Microbiology* **145 (Pt 4)**:767-779.
- Ruttenberg, M. A. und B. Mach. 1966. Studies on amino acid substitution in the biosynthesis of the antibiotic polypeptide tyrocidine. *Biochemistry* **5**:2864-2869.
- Salemi, M. und andere. 2003. *The phylogenetic handbook*. Cambridge University Press, Cambridge (UK).
- Sambrook, J. und D. W. Russell. 2001. *Molecular Cloning. A laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
- Sandmann, A., J. Dickschat, H. Jenke-Kodama, B. Kunze, E. Dittmann und R. Müller. 2007. A Type II polyketide synthase from the gram-negative *Bacterium Stigmatella aurantiaca* is involved in Aurachin alkaloid biosynthesis. *Angew Chem Int Ed Engl* **46**:2712-2716.
- Santoyo, G. und D. Romero. 2005. Gene conversion and concerted evolution in bacterial genomes. *FEMS Microbiol Rev* **29**:169-183.

- Sarkanen, K. V. und C. H. Ludwig. 1971. Lignins: occurrence, formation, structure und reactions. Wiley Interscience, New York.
- Saxena, P., G. Yadav, D. Mohanty und R. S. Gokhale. 2003. A new family of type III polyketide synthases in *Mycobacterium tuberculosis*. *J Biol Chem* **278**:44780-44790.
- Schmidt, H. A., K. Strimmer, M. Vingron und A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**:502-504.
- Schneider, R., B. Brors, M. Massow und H. Weiss. 1997. Mitochondrial fatty acid synthesis: a relic of endosymbiotic origin and a specialized means for respiration. *FEBS Lett* **407**:249-252.
- Schneider, T. L., B. Shen und C. T. Walsh. 2003. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42**:9722-9730.
- Schröder, J. 1999. The chalcone/stilbene synthase-type family of condensing enzymes. Pp. 749-771 in U. Sankawa, ed. *Polyketides and other secondary metabolites including fatty acids and their derivatives*. Elsevier, Oxford.
- Schwecke, T., J. F. Aparicio, I. Molnar, A. König, L. E. Khaw, S. F. Haydock, M. Oliynyk, P. Caffrey, J. Cortes, J. B. Lester und et al. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc Natl Acad Sci U S A* **92**:7839-7843.
- Schweizer, E. und J. Hofmann. 2004. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol. Mol. Biol. Rev.* **68**:501-517, table of contents.
- Schweizer, E., K. Werkmeister und M. K. Jain. 1978. Fatty acid biosynthesis in yeast. *Mol Cell Biochem* **21**:95-107.
- Scott, A. F., P. Heath, S. Trusko, S. H. Boyer, W. Prass, M. Goodman, J. Czelusniak, L. Y. Chang und J. L. Slightom. 1984. The sequence of the gorilla fetal globin genes: evidence for multiple gene conversions in human evolution. *Mol Biol Evol* **1**:371-389.
- Serre, L., E. C. Verbree, Z. Dauter, A. R. Stuitje und Z. S. Derewenda. 1995. The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase at 1.5-Å resolution. Crystal structure of a fatty acid synthase component. *J Biol Chem* **270**:12961-12964.
- Seshime, Y., P. R. Juvvadi, I. Fujii und K. Kitamoto. 2005. Discovery of a novel superfamily of type III polyketide synthases in *Aspergillus oryzae*. *Biochem Biophys Res Commun* **331**:253-260.
- Shen, B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* **7**:285-295.
- Shen, B., L. Du, C. Sanchez, D. J. Edwards, M. Chen und J. M. Murrell. 2001. The biosynthetic gene cluster for the anticancer drug bleomycin from *Streptomyces verticillus* ATCC15003 as a model for hybrid peptide-polyketide natural product biosynthesis. *J Ind Microbiol Biotechnol* **27**:378-385.
- Shen, B. und C. R. Hutchinson. 1993. Enzymatic synthesis of a bacterial polyketide from acetyl and malonyl coenzyme A. *Science* **262**:1535-1540.
- Sherman, D. H. 2005. The Lego-ization of polyketide biosynthesis. *Nat Biotechnol* **23**:1083-1084.
- Shoolingin-Jordan, P. M. und I. D. G. Campuzano. 1999. Biosynthesis of 6-Methylsalicylic Acid. Pp. 345-365 in U. Sankawa, ed. *Polyketides and other secondary metabolites including fatty acids and their derivatives*. Elsevier, Oxford.
- Sieber, S. A. und M. A. Marahiel. 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* **105**:715-738.
- Siggaard-Andersen, M. 1993. Conserved residues in condensing enzyme domains of fatty acid synthases and related sequences. *Protein Seq Data Anal* **5**:325-335.
- Silakowski, B., H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blocker, G. Hofle, S. Beyer und R. Müller. 1999. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J Biol Chem* **274**:37391-37399.
- Smith, S. 2006. Structural biology. Architectural options for a fatty acid synthase. *Science* **311**:1251-1252.
- Smith, S., A. Witkowski und A. K. Joshi. 2003. Structural and functional organization of the animal fatty acid synthase. *Prog Lipid Res* **42**:289-317.
- Stachelhaus, T., A. Huser und M. A. Marahiel. 1996. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chem Biol* **3**:913-921.
- Stachelhaus, T., H. D. Mootz und M. A. Marahiel. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem Biol* **6**:493-505.
- Stachelhaus, T., A. Schneider und M. A. Marahiel. 1995. Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains. *Science* **269**:69-72.

- Stassi, D. L., S. J. Kakavas, K. A. Reynolds, G. Gunawardana, S. Swanson, D. Zeidner, M. Jackson, H. Liu, A. Buko und L. Katz. 1998. Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proc Natl Acad Sci U S A* **95**:7305-7309.
- Staunton, J. und K. J. Weissman. 2001. Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**:380-416.
- Staunton, J. und B. Wilkinson. 1998. The biosynthesis of aliphatic polyketides. *Topics Curr Chem*. **195**:49-92.
- Staunton, J. und B. Wilkinson. 2001. Combinatorial biosynthesis of polyketides and nonribosomal peptides. *Curr Opin Chem Biol* **5**:159-164.
- Stechmann, A. und T. Cavalier-Smith. 2002. Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**:89-91.
- Steenkamp, E. T., J. Wright und S. L. Baldauf. 2006. The protistan origins of animals and fungi. *Mol Biol Evol* **23**:93-106.
- Sthapit, B., T. J. Oh, R. Lamichhane, K. Liou, H. C. Lee, C. G. Kim und J. K. Sohng. 2004. Neocarzinostatin naphthoate synthase: an unique iterative type I PKS from neocarzinostatin producer *Streptomyces carzinostaticus*. *FEBS Lett* **566**:201-206.
- Stinear, T. P., A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay und S. T. Cole. 2004. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A* **101**:1345-1349.
- Stoops, J. K. und S. J. Wakil. 1981. Animal fatty acid synthetase. A novel arrangement of the beta-ketoacyl synthetase sites comprising domains of the two subunits. *J Biol Chem* **256**:5128-5133.
- Summers, R. G., A. Ali, B. Shen, W. A. Wessel und C. R. Hutchinson. 1995. Malonyl-coenzyme A:acyl carrier protein acyltransferase of *Streptomyces glaucescens*: a possible link between fatty acid and polyketide biosynthesis. *Biochemistry* **34**:9389-9402.
- Sun, Y., X. Zhou, H. Dong, G. Tu, M. Wang, B. Wang und Z. Deng. 2003. A complete gene cluster from *Streptomyces nanchangensis* NS3226 encoding biosynthesis of the polyether ionophore nanchangmycin. *Chem Biol* **10**:431-441.
- Sutherland, J. M., M. Herdman und W. D. P. Stewart. 1979. Akinetes of the cyanobacterium *Nostoc* PCC 7524: macromolecular composition, structure and control of differentiation. *J Gen Microbiol* **115**:273-287.
- Swofford, D. 2002. PAUP* 4.0: Phylogenetic analysis using parsimony. Sinauer Associates, Sunderland, MA.
- Szekeres, M., K. Nemeth, Z. Koncz-Kalman, J. Mathur, A. Kauschmann, T. Altmann, G. P. Redei, F. Nagy, J. Schell und C. Koncz. 1996. Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in *Arabidopsis*. *Cell* **85**:171-182.
- Tanabe, Y., K. Kaya und M. M. Watanabe. 2004. Evidence for recombination in the microcystin synthetase (mcy) genes of toxic cyanobacteria *Microcystis* spp. *J Mol Evol* **58**:633-641.
- Tang, G. L., Y. Q. Cheng und B. Shen. 2004. Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. *Chem Biol* **11**:33-45.
- Tekaia, F. und B. Dujon. 1999. Pervasiveness of gene conservation and persistence of duplicates in cellular genomes. *J Mol Evol* **49**:591-600.
- Theißen, G. 2002. Orthology: secret life of genes. *Nature* **415**:741.
- Thiel, T. und C. P. Wolk. 1983. Metabolic activities of isolated akinetes of the cyanobacterium *Nostoc spongiaeforme*. *J Bacteriol* **156**:369-374.
- Thompson, J. D., D. G. Higgins und T. J. Gibson. 1994. ClustalW: Improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. **22**.
- Toomey, R. E. und S. J. Wakil. 1966. Studies on the mechanism of fatty acid synthesis. XV. Preparation and general properties of beta-ketoacyl acyl carrier protein reductase from *Escherichia coli*. *Biochim Biophys Acta* **116**:189-197.
- Tropf, S., B. Karcher, G. Schroder und J. Schroder. 1995. Reaction mechanisms of homodimeric plant polyketide synthase (stilbenes and chalcone synthase). A single active site for the condensing reaction is sufficient for synthesis of stilbenes, chalcones und 6'-deoxychalcones. *J Biol Chem* **270**:7922-7928.

- Tropf, S., T. Lanz, S. A. Rensing, J. Schroder und G. Schroder. 1994. Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. *J Mol Evol* **38**:610-618.
- Tuan, J. S., J. M. Weber, M. J. Staver, J. O. Leung, S. Donadio und L. Katz. 1990. Cloning of genes involved in erythromycin biosynthesis from *Saccharopolyspora erythraea* using a novel actinomycete-*Escherichia coli* cosmid. *Gene* **90**:21-29.
- Turgay, K., M. Krause und M. A. Marahiel. 1992. Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes. *Mol Microbiol* **6**:2743-2744.
- Ueda, K., K. M. Kim, T. Beppu und S. Horinouchi. 1995. Overexpression of a gene cluster encoding a chalcone synthase-like protein confers redbrown pigment production in *Streptomyces griseus*. *J Antibiot (Tokyo)* **48**:638-646.
- Vagelos, P. R., A. W. Alberts und P. W. Majerus. 1969. Mechanism of saturated fatty acid biosynthesis in *Escherichia coli*. *Methods Enzymol* **14**:39-42.
- Valway, S. E., M. P. Sanchez, T. F. Shinnick, I. Orme, T. Agerton, D. Hoy, J. S. Jones, H. Westmoreland und I. M. Onorato. 1998. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis*. *N Engl J Med* **338**:633-639.
- Vandamme, A.-M. 2003. Kapitel 1: Basic concepts of molecular evolution *in* M. Salemi und A.-M. Vandamme, eds. *The phylogenetic handbook*. Cambridge University Press, Cambridge (UK).
- Vining, L. C. 1992. Roles of secondary metabolites from microbes. Pp. 184-198. Ciba Foundation Symposia.
- Vissa, V. D. und P. J. Brennan. 2001. The genome of *Mycobacterium leprae*: a minimal mycobacterial gene set. *Genome Biol* **2**:REVIEWS1023.
- Vogel, J., I. M. Axmann, H. Herzel und W. R. Hess. 2003. Experimental and computational analysis of transcriptional start sites in the cyanobacterium *Prochlorococcus* MED4. *Nucleic Acids Res* **31**:2890-2899.
- Volodin, A. A., O. N. Voloshin und R. D. Camerini-Otero. 2005. Homologous recombination and RecA protein: towards a new generation of tools for genome manipulations. *Trends Biotechnol* **23**:97-102.
- Wakil, S. J. 1961. Mechanism of fatty acid synthesis. *J Lipid Res* **2**:1-24.
- Wallis, J. G., J. L. Watts und J. Browse. 2002. Polyunsaturated fatty acid synthesis: what will they think of next? *Trends Biochem Sci* **27**:467.
- Watanabe, K., C. Ishikawa, I. Ohtsuka, M. Kamata, M. Tomita, K. Yazawa und H. Muramatsu. 1997. Lipid and fatty acid compositions of a novel docosahexaenoic acid-producing marine bacterium. *Lipids* **32**:975-978.
- Weber, G., K. Schorgendorfer, E. Schneider-Scherzer und E. Leitner. 1994. The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr Genet* **26**:120-125.
- Weber, T. und M. A. Marahiel. 2001. Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure* **9**:R3-9.
- Weiss, E., L. Golden, R. Zakut, A. Mellor, K. Fahrner, S. Kvist und R. A. Flavell. 1983a. The DNA sequence of the H-2kb gene: evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. *Embo J* **2**:453-462.
- Weiss, E. H., A. Mellor, L. Golden, K. Fahrner, E. Simpson, J. Hurst und R. A. Flavell. 1983b. The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion-like events. *Nature* **301**:671-674.
- Weissman, K. J. 2004. Polyketide biosynthesis: understanding and exploiting modularity. *Philos Transact A Math Phys Eng Sci* **362**:2671-2690.
- Weissman, K. J. und P. F. Leadlay. 2005. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**:925-936.
- Weissman, K. J., M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton und P. F. Leadlay. 1997. The molecular basis of Celmer's rules: the stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase. *Biochemistry* **36**:13849-13855.
- Weitnauer, G., A. Muhlenweg, A. Trefzer, D. Hoffmeister, R. D. Sussmuth, G. Jung, K. Welzel, A. Vente, U. Girreser und A. Bechthold. 2001. Biosynthesis of the orthosomycin antibiotic avilamycin A: deductions from the molecular analysis of the avi biosynthetic gene cluster of *Streptomyces viridochromogenes* Tu57 and production of new antibiotics. *Chem Biol* **8**:569-581.

- Welker, M. und H. von Döhren. 2006. Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS Microbiol Rev* **30**:530-563.
- Wenzel, S. C., B. Kunze, G. Hofle, B. Silakowski, M. Scharfe, H. Blocker und R. Müller. 2005. Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *Chembiochem* **6**:375-385.
- Wenzel, S. C. und R. Müller. 2005. Formation of novel secondary metab. *Curr Opin Chem Biol* **9**:447-458.
- Whitton, B. A. und M. Potts. 2000. Pp. 1-11 in B. A. Whitton und M. Potts, eds. *The ecology of cyanobacteria*. Kluwer, Dordrecht.
- Wiener, P., S. Egan, A. S. Huddleston und E. M. Wellington. 1998. Evidence for transfer of antibiotic-resistance genes in soil populations of streptomycetes. *Mol Ecol* **7**:1205-1216.
- Wiesmann, K. E., J. Cortes, M. J. Brown, A. L. Cutter, J. Staunton und P. F. Leadlay. 1995. Polyketide synthesis in vitro on a modular polyketide synthase. *Chem Biol* **9**:583-589.
- Wilkinson, B., G. Foster, B. A. Rudd, N. L. Taylor, A. P. Blackaby, P. J. Sidebottom, D. J. Cooper, M. J. Dawson, A. D. Buss, S. Gaisser, I. U. Bohm, C. J. Rowe, J. Cortes, P. F. Leadlay und J. Staunton. 2000. Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. *Chem Biol* **7**:111-117.
- Wilkinson, B. und J. Micklefield. 2007. Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* **3**:379-386.
- Wilkinson, C. J., E. J. Frost, J. Staunton und P. F. Leadlay. 2001. Chain initiation on the soraphen-producing modular polyketide synthase from *Sorangium cellulosum*. *Chem Biol* **8**:1197-1208.
- Williams, D. H., M. J. Stone, P. R. Hauck und S. K. Rahman. 1989. Why are secondary metabolites (natural products) biosynthesized? *J Nat Prod* **52**:1189-1208.
- Winz, R. A. und I. T. Baldwin. 2001. Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. IV. Insect-Induced ethylene reduces jasmonate-induced nicotine accumulation by regulating putrescine N-methyltransferase transcripts. *Plant Physiol* **125**:2189-2202.
- Witkowski, A., A. K. Joshi, V. S. Rangan, A. M. Falick, H. E. Witkowska und S. Smith. 1999. Dibromopropanone cross-linking of the phosphopantetheine and active-site cysteine thiols of the animal fatty acid synthase can occur both inter- and intrasubunit. Reevaluation of the side-by-side, antiparallel subunit model. *J Biol Chem* **274**:11557-11563.
- Woese, C. R. 1987. Bacterial evolution. *Microbiol Rev* **51**:221-271.
- Woese, C. R. 2000. Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci U S A* **97**:8392-8396.
- Woese, C. R., O. Kandler und M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria und Eucarya. *Proc Natl Acad Sci U S A* **87**:4576-4579.
- Wolf, Y. I., A. S. Kondrashov und E. V. Koonin. 2000. Interkingdom gene fusions. *Genome Biol* **1**:RESEARCH0013.
- Wolk, C. P. 2000. Heterocyst formation in *Anabaena*. Pp. 83-104 in Y. V. Brun und L. J. Shimkets, eds. *Prokaryotic development*. American Society of Microbiology, Washington, D. C.
- Wolk, C. P., A. Ernst und J. Elhai. 1994. Heterocyst metabolism and development. Pp. 769-823 in D. A. Bryant, ed. *The molecular biology of cyanobacteria*. Kluwer, Boston.
- Wolk, C. P., A. Vonshak, P. Kehoe und J. Elhai. 1984. Construction of shuttle vectors capable of conjugative transfer from *Escherichia coli* to nitrogen-fixing filamentous cyanobacteria. *Proc Natl Acad Sci U S A* **81**:1561-1565.
- Wu, J., T. J. Zaleski, C. Valenzano, C. Khosla und D. E. Cane. 2005. Polyketide double bond biosynthesis. Mechanistic analysis of the dehydratase-containing module 2 of the picromycin/methymycin polyketide synthase. *J Am Chem Soc* **127**:17393-17404.
- Wu, K., L. Chung, W. P. Revill, L. Katz und C. D. Reeves. 2000a. The FK520 gene cluster of *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) contains genes for biosynthesis of unusual polyketide extender units. *Gene* **251**:81-90.
- Wu, N., F. Kudo, D. E. Cane und C. Khosla. 2000b. Analysis of the molecular recognition features of individual modules derived from the erythromycin polyketide synthase. *J Am Chem Soc* **122**:4847-4852.
- Xia, X. und Z. Xie. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* **92**:371-373.

- Xu, P., G. Widmer, Y. Wang, L. S. Ozaki, J. M. Alves, M. G. Serrano, D. Puiu, P. Manque, D. Akiyoshi, A. J. Mackey, W. R. Pearson, P. H. Dear, A. T. Bankier, D. L. Peterson, M. S. Abrahamsen, V. Kapur, S. Tzipori und G. A. Buck. 2004. The genome of *Cryptosporidium hominis*. *Nature* **431**:1107-1112.
- Yadav, G., R. S. Gokhale und D. Mohanty. 2003. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**:335-363.
- Yang, D., Y. Oyaizu, H. Oyaizu, G. J. Olsen und C. R. Woese. 1985. Mitochondrial origins. *Proc Natl Acad Sci U S A* **82**:4443-4447.
- Yang, Z., N. Goldman und A. Friday. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. *Mol Biol Evol* **11**:316-324.
- Yasuno, R., P. von Wettstein-Knowles und H. Wada. 2004. Identification and molecular characterization of the beta-ketoacyl-[acyl carrier protein] synthase component of the Arabidopsis mitochondrial fatty acid synthase. *J Biol Chem* **279**:8242-8251.
- Yazawa, K. 1996. Production of eicosapentaenoic acid from marine bacteria. *Lipids* **31 Suppl**:S297-300.
- Zähner, H. 1979. What are secondary metabolites? *Folia Microbiol* **24**:435-443.
- Zähner, H. und T. Anke. 1983. Evolution of secondary pathways. Pp. 153-171 in J. W. Bennet und E. Ciegler, eds. *Differentiation and secondary metabolism in fungi*. Marcel Dekker, New York.
- Zenk, M. H. 1991. Chasing the enzymes of secondary metabolism: plant cell cultures as a pot of gold. *Phytochemistry* **30**:3861-3863.
- Zhang, J., H. F. Rosenberg und M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* **95**:3708-3713.
- Zhang, J. R., J. M. Hardham, A. G. Barbour und S. J. Norris. 1997. Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**:275-285.
- Zhang, L., A. K. Joshi, J. Hofmann, E. Schweizer und S. Smith. 2005. Cloning, expression und characterization of the human mitochondrial beta-ketoacyl synthase. Complementation of the yeast CEM1 knock-out strain. *J Biol Chem* **280**:12422-12429.
- Zhou, P., G. Florova und K. A. Reynolds. 1999. Polyketide synthase acyl carrier protein (ACP) as a substrate and a catalyst for malonyl ACP biosynthesis. *Chem Biol* **6**:577-584.
- Zhu, G., M. J. LaGier, F. Stejskal, J. J. Millership, X. Cai und J. S. Keithly. 2002. *Cryptosporidium parvum*: the first protist known to encode a putative polyketide synthase. *Gene* **298**:79-89.

Abkürzungsverzeichnis

AA-Medium	Allan-Arnon-Medium
ACP	Acyl-Trägerprotein (<i>acyl carrier protein</i>)
AcT	Acyltransferase, spezialisierte Form in Pilzen
A-Domäne	Adenylierungsdomäne
AMP	Adeninmononucleotid
AT	Acyltransferase
ATP	Adenosin-Triphosphat
bp	Basenpaare
C-Domäne	Kondensationsdomäne (<i>condensation domain</i>)
CHS	Chalcon-Synthase
CLF	Kettenlängen-Faktor (<i>chain length factor</i>)
CMN-Gruppe	Bakteriengruppe, die Corynebakterien, Mycobakterien und Nocardien umfasst
CoA	Coenzym A
Da	Dalton
DH	Dehydratase
DNA	Desoxyribonucleinsäure
DTT	Dithiotreitol
E-Domäne	Epimerisierungsdomäne
EMSA	Elektrophoretische Mobilitätsverschiebungs-Test (<i>electrophoretic mobility shift assay</i>)
ER	Enoylreductase
FAS	Fettsäure-Synthase (<i>fatty acid synthase</i>)
HGT	horizontaler Gentransfer
HPLC	Hochauflösende Flüssigkeitschromatographie (<i>high performance liquid chromatography</i>)
kbp	Kilobasenpaare
KR	Ketoreductase
KS	Ketosynthase
LB-Medium	Luria-Bertani-Medium
MALDI	Matrix-unterstützt Laser-Desorptions-Ionisation (<i>matrix-assisted laser desorption ionisation</i>)
MAT	Malonyl-/Acetyl-Transferase
MCMC	Markov-Ketten-Monte-Carlo(-Analyse)
MDa	Megadalton
M-Domäne	Methylierungsdomäne
MOPS	3-Morpholino-Propansulfonsäure
MP	<i>Maximum parsimony</i>
NADP	Nicotinamidadenindinucleotid, oxidierte Form
NADPH	Nicotinamidadenindinucleotid, reduzierte Form
NJ	<i>Neighbour joining</i>
NRPS	Nichtribosomale Peptid-Synthasen
Ox-Domäne	Oxidase-Domäne

PAGE	Polyacrylamid-Gelelektrophorese
PCP	Peptidyl-Trägerprotein (<i>peptidyl carrier protein</i>)
PCR	Polymerase-Kettenreaktion (<i>polymerase chain reaction</i>)
PKS	Polyketid-Synthase
PPI	Pyrophosphat
PUFA	vielfach ungesättigte Fettsäuren (<i>polyunsaturated fatty acids</i>)
RNA	Ribonucleinsäure
SAM	S-Adenosylmethionin
SDS	Natrium-Dodecylsulfat
SPE	Festphasen-Extraktion (<i>solid phase extraction</i>)
STS	Stilben-Synthase
TBS	Tris-gepufferte Salzlösung (<i>tris buffered saline</i>)
TE	Thioesterase
TE-Puffer	Tris-EDTA-Puffer
TOF	<i>time of flight</i>
tRNA	Transport-Ribonucleinsäure

Danksagungen

Zum Gelingen dieser Arbeit haben viele Menschen beigetragen – sowohl in fachlicher wie auch in privater Hinsicht. Ihnen allen spreche ich meinen herzlichen Dank für ihre Unterstützung dieses Projekts aus!

Konkrete Aufzählungen bergen natürlich die Gefahr, einen oder mehrere Namen zu vergessen. Ich gehe dieses kleine Wagnis dennoch ein und bedanke mich besonders bei

Prof. Elke Dittmann und Prof. Thomas Börner für die Aufnahme in ihre Arbeitsgruppen, die Themenstellung und Diskussionsbereitschaft,

PD Dr. Hans von Döhren, Dr. Torsten Schwecke, Dr. Martin Welker und Kathrin Schubert, alle von der TU Berlin, für die Unterstützung bei einem Teil der praktischen Laborarbeiten und die Durchführung der massenspektrometrischen Analysen,

Prof. Rolf Müller (Universität des Saarlandes), mit dessen Gruppe einige erfolgreiche Kooperationen stattfanden,

der gesamten Arbeitsgruppe Genetik des Instituts für Biologie der HU Berlin für die gute Arbeitsatmosphäre und einfach all die schönen Jahre, die ich dort verbringen konnte, wobei Brita Fiedler, Carola Emanuel, Jan Kehr, Kristina Kühn, Nadine Ziemert, Yvonne Zilliges und Katrin Hinrichs besonders hervorgehoben seien,

der Deutschen Forschungsgemeinschaft für den Großteil der Finanzierung des Projekts über das Schwerpunktprogramm 1152 zur „Evolution metabolischer Diversität“,

und schließlich...

....meiner Familie, die trotz mancher Verzögerung die Hoffnung nicht aufgab und mich auf jede Weise unterstützte.

Und nicht zuletzt:

Ein extragroßes **Dankeschön** an meine Frau Akiko, die mein Leben schöner macht.

Publikationen

Wilson DN, Blaha G, Connell SR, Ivanov PV, **Jenke H**, Stelzl U, Teraoka Y, Nierhaus KH (2002). Protein synthesis at atomic resolution: mechanistics of translation in the light of highly resolved structures for the ribosome. *Curr. Protein Pept. Sci.*, Feb 3(1): 1-53.

Jenke-Kodama H, Sandmann A, Müller R, Dittmann E (2005). Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.*, Oct 22 (10): 2027-2039.

Jenke-Kodama H, Dittmann E (2005). Combinatorial polyketide biosynthesis at higher stage. *Molecular Systems Biology*, DOI: 10.1038/msb4100033.

Gross F, Luniak N, Perlova O, Gaitatzis N, **Jenke-Kodama H**, Gerth K, Gottschalk D, Dittmann E, Müller R (2006). Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.*, 185(1): 28-38.

Jenke-Kodama H, Börner T, Dittmann E (2006). Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. *PLoS Comput. Biol.*, Oct 6; 2(10), DOI: 10.1371/journal.pcbi.0020132.

Sandmann A, Dickschat J, **Jenke-Kodama H**, Dittmann E, Müller R (2007). A type II polyketide synthase from the gram-negative bacterium *Stigmatella aurantiaca* is involved in Aurachin alkaloid biosynthesis. *Angew. Chem. Int. Ed. Engl.*, 46: 2712-2716.

Jenke-Kodama H, Müller R, Dittmann E (2007). Evolutionary mechanisms underlying secondary metabolite diversity. *Novartis Review Series* (in Druck).

Traitcheva N, He J, Busch B, **Jenke-Kodama H**, Dittmann E, Hertweck C (2007). Non-colinear polyketide biosynthesis in the Aureothin and Neoareothin pathways: An evolutionary perspective. *ChemBioChem* (in Druck).

Guljamow A, **Jenke-Kodama H**, Saumweber H, Quillardet P, Frangeul L, Castets AM, Bouchier C, Tandeau de Marsac N, Dittmann E (2007). Horizontal gene transfer of two cytoskeletal elements from a eukaryote to a cyanobacterium. *Curr. Biol.* (in Druck).

Brachmann AO, Joyce SA, **Jenke-Kodama H**, Schwär G, Clarke DJ, Bode HB (2007). Anthraquinone biosynthesis in *Photorhabdus luminescens*. *ChemBioChem* (in Druck).

Erklärung

Ich versichere hiermit, dass ich die vorliegende Dissertation im Fach Biologie selbstständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die dem Verfahren zugrunde liegende Promotionsordnung ist mir bekannt.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.